Seeing the Forest Through the Trees: The Distribution and Properties of Dense Molecular Gas in the Milky Way Galaxy

by

Timothy P. Ellsworth-Bowers

B.S., University of Arizona, 2002

B.S.S.E., University of Arizona, 2002

M.S., University of Colorado, 2010

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Astrophysical and Planetary Sciences

2014

This thesis entitled:
Seeing the Forest Through the Trees:
The Distribution and Properties of Dense Molecular Gas in the Milky Way Galaxy
written by Timothy P. Ellsworth-Bowers
has been approved for the Department of Astrophysical and Planetary Sciences

Prof. Jason Glenn

Prof. John Bally

Prof. Jack Burns

Prof. Albin Gasiewski

Prof. David Brain

Date

The final copy of this thesis has been examined by the signatories, and we find that both the content and the form meet acceptable presentation standards of scholarly work in the above mentioned discipline.

Ellsworth-Bowers, Timothy P. (Ph.D., Astrophysics)

Seeing the Forest Through the Trees:

The Distribution and Properties of Dense Molecular Gas in the Milky Way Galaxy

Thesis directed by Prof. Jason Glenn

The Milky Way Galaxy serves as a vast laboratory for studying the dynamics and evolution of the dense interstellar medium and the processes of and surrounding massive star formation. From our vantage point within the Galactic plane, however, it has been extremely difficult to construct a coherent picture of Galactic structure; we cannot see the forest for the trees. The principal difficulties in studying the structure of the Galactic disk have been obscuration by the ubiquitous dust and molecular gas and confusion between objects along a line of sight. Recent technological advances have led to large-scale blind surveys of the Galactic plane at (sub-)millimeter wavelengths, where Galactic dust is generally optically thin, and have opened a new avenue for studying the forest.

The Bolocam Galactic Plane Survey (BGPS) observed over 190 deg^2 of the Galactic plane in dust continuum emission near $\lambda = 1.1$ mm, producing a catalog of over 8,000 dense molecular cloud structures across a wide swath of the Galactic disk. Deriving the spatial distribution and physical properties of these objects requires knowledge of distance, a component lacking in the data themselves. This thesis presents a generalized Bayesian probabilistic distance estimation method for dense molecular cloud structures, and demonstrates it with the BGPS data set. Distance probability density functions (DPDFs) are computed from kinematic distance likelihoods (which may be double-peaked for objects in the inner Galaxy) and an expandable suite of prior information to produce a comprehensive tally of our knowledge (and ignorance) of the distances to dense molecular cloud structures. As part of the DPDF formalism, this thesis derives several prior DPDFs for resolving the kinematic distance ambiguity in the inner Galaxy.

From the collection of posterior DPDFs, a set of objects with well-constrained distance estimates is produced for deriving Galactic structure and the physical properties of dense molecular cloud structures. This distance catalog of 1,802 objects across the Galactic plane represents the first large-scale analysis of clump-scale objects in a variety of Galactic environments. The Galactocentric positions of these objects begin to trace out the spiral structure of the Milky Way, and suggest that dense molecular gas settles nearer the Galactic midplane than tracers of less-dense gas such as CO. Physical properties computed from the DPDFs reveal that BGPS objects trace a continuum of scales within giant molecular clouds, and extend the scaling relationships known as Larson's Laws to lower-mass substructures. The results presented here represent the first step on the road to seeing the molecular content of the Milky Way as a forest rather than individual nearby trees.

Dedication

To Esther, Eugene, Anna, and Emily. Your patience, support, kindness and love have made this journey possible.

Acknowledgements

I gratefully acknowledge my advisor, Jason Glenn, for always being enthusiastic about my work and for being graciously understanding and supportive during the challenging times of graduate school. I appreciate his willingness to bring me on early to a large-scale **Herschel** extragalactic project, and his resourcefulness when I decided to study a Galaxy a little closer to home. His excitement about my work has helped pull me through to the end of this.

Thank you to Jesse Lord, whose friendship and support have helped me keep my feet on the ground amidst the swirls of graduate school. Also to the other members of our incoming class, who helped me through the first year of classes to remember how to be a student again.

For help with the work contained in this thesis, I am foremost grateful to Erik Rosolowsky, who provided the kernels for both the mathematical and computer-code frameworks upon which I built this work, and also many "blue sky" ideas to chase down. Thank you to Adam Ginsburg, who helped me understand the BGPS data and the BGPS collaboration. Other member of the BGPS group helping with this opus were John Bally, Cara Battersby, Neal Evans, Yancy Shirley, and Brian Svoboda. Thanks also to Yancy for hauling me up the SMT to use it for actual observing.

But above all, thank you to all of my family who have supported me and enabled me to tackle this dream, especially Esther. I couldn't have done it without you.

Contents

Chapter

1 Introduction 1

1.1 Preface 1

1.2 Star Formation in the Milky Way 2

1.2.1 Overview of Star Formation 2

1.2.2 The Milky Way: Seeing the Forest Through the Trees 4

1.3 Observations of Star Formation Tracers 8

1.3.1 How Do We Observe Star Forming Regions? 8

1.3.2 The Bolocam Galactic Plane Survey 9

1.4 Scope of Work 12

1.4.1 A Probabilistic Distance Estimation Method 12

1.4.2 Galactic Structure and the Physical Properties of Dense Molecular Cloud Structures 13

1.5 Organization 14

2 A Mid-Infrared Kinematic Distance Discrimination Method 15

2.1 Introduction 16

2.2 Data Sets 19

2.2.1 The Bolocam Galactic Plane Survey 19

2.2.2 Spectroscopic Follow-Up of BGPS Sources 20

2.2.3 The **Spitzer** GLIMPSE Survey . . . 22

2.3 Distance Probability Density Functions . . . 22

2.3.1 Approach and Utility . . . 22

2.3.2 Extracting a Distance from the DPDF . . . 23

2.3.3 Using DPDFs to Estimate Physical Parameters . . . 24

2.3.4 Kinematic Distance DPDFs . . . 25

2.3.5 Prior DPDFs for Kinematic Distance Discrimination . . . 27

2.4 Infrared-Millimeter Morphological Matching . . . 27

2.4.1 Creation of Synthetic 8-μm Images . . . 28

2.4.2 Processing of GLIMPSE 8-μm Images . . . 34

2.4.3 Morphological Matching . . . 37

2.5 Results . . . 40

2.5.1 EMAF-Selected Molecular Cloud Clumps . . . 40

2.5.2 Source KDA Resolutions . . . 46

2.5.3 Distance Comparisons with Other Studies . . . 53

2.6 Discussion . . . 58

2.6.1 Kinematic Distance Discrimination . . . 58

2.6.2 EMAF vs. IRDC . . . 63

2.6.3 Implications for Galactic Structure . . . 64

2.7 Summary . . . 67

2.8 Appendices . . . 69

2.8.1 Computing a Band-Averaged Dust Opacity: **Spitzer** IRAC Band 4 . . . 69

2.8.2 **Spitzer** IRAC Scattering Correction Factors . . . 72

2.8.3 The Vertical Solar Offset and Converting $(\ell, b, d_\odot)$ into $(R_{\rm gal}, \phi, z)$ . . . 75

3 Distance Catalog Expansion Using Kinematic Isolation of Dense Molecular Cloud Structures With ^{13}CO(1-0) 77

3.1 Introduction 78
3.2 Data 81
3.2.1 The Bolocam Galactic Plane Survey 81
3.2.2 Molecular Line Spectroscopic Observations 82
3.2.3 The Galactic Ring Survey 83
3.2.4 The `distance-omnibus` Code Base 83
3.3 DPDFs and Galactic Kinematics 83
3.3.1 Overview of the DPDF Formalism 83
3.3.2 Galactic Kinematics 85
3.4 Molecular Cloud Clump ^{13}CO Velocity Extraction 89
3.4.1 Comparison of GRS ^{13}CO Spectra with Dense Gas Tracers 89
3.4.2 Morphological Spectrum Extraction Technique 90
3.4.3 The ^{13}CO Kinematic Distance Likelihood Function 96
3.5 Catalog-Based Prior DPDFs 97
3.5.1 Associating Molecular Cloud Structures with Literature Catalog Objects . . . 97
3.5.2 Trigonometric Parallax Measurements 101
3.5.3 H II Region KDA Resolutions 102
3.6 Results 105
3.6.1 Kinematics 105
3.6.2 Prior DPDFs 110
3.6.3 Analysis of "Well-Constrained" Distance Estimates 114
3.6.4 The V2 Distance Catalog and Posterior DPDFs 115
3.6.5 Galactocentric Positions 121
3.7 Discussion 125
3.7.1 ^{13}CO as Molecular Cloud Clump Tracer 125
3.7.2 Inter-Comparison of Prior DPDF Methods 127
3.7.3 A Representative Sample? 129

3.8 Summary 131

4 Physical Properties and Mass Functions of Dense Molecular Cloud Structures 134

4.1 Introduction 135

4.2 Data 137

4.3 Method 139

4.3.1 Computing Physical Properties Using the DPDF 139

4.3.2 Constructing Mass Distributions and Estimating Completeness 142

4.3.3 Fitting a Functional Form to the Mass Distribution 146

4.4 Results 149

4.4.1 Physical Properties of BGPS Sources 149

4.4.2 Mass Function Fits for Dense Molecular Cloud Structures 157

4.4.3 Mass Functions Based on Astrophysical Cuts 165

4.4.4 Larson's Scaling Relationships 174

4.4.5 Distribution of Dense Gas in the Galactic Plane 177

4.5 Discussion 181

4.5.1 What is a BGPS source? 181

4.5.2 Larson's Laws & The Continuum of Structure 183

4.5.3 Mass Distributions and Mass Functions 186

4.5.4 Estimating the Dense Gas Mass Fraction 189

4.6 Summary 192

5 Concluding Thoughts 195

5.1 Summary 195

5.2 Future Work 198

5.2.1 Additional Prior DPDFs 198

5.2.2 Application of the DPDF Formalism to **e.g.**, Hi-GAL and CCAT 201

Bibliography 203

Appendix

A The `distance-omnibus` Code Base 210

A.1 Motivation 210

A.2 Modular Framework 211

A.3 Public Access and Future Development 212

Tables

Table

2.1 Spectroscopic Follow-up Observations of Inner Galaxy BGPS Sources 19

2.2 Comparison of Hyperion Model Parameters 31

2.3 Observed & Derived Properties of EMAF-selected BGPS Molecular Cloud Clumps . 44

2.4 Maser Sources for Distance Comparison 50

2.5 Effect of Dust Temperature on KDA Resolution 58

2.6 Computed $\langle\kappa\rangle_{\mathrm{band}}$ for IRAC Band 4 71

3.1 Galactic Rotation Curve Parameters 85

3.2 H II Region KDA Resolutions from the HRDS 103

3.3 Number of BGPS V2 Sources for Each Velocity Tracer 106

3.4 KDA Resolutions for BGPS V2 Sources 116

3.5 BGPS V2 Distance Catalog 117

3.6 Application of DPDFs to BGPS V2 Sources 126

3.7 Well-Constrained Fraction for Overlapping Priors 127

4.1 Physical Properties of BGPS V2 Sources 150

4.2 Mass Distribution Functional Fit Parameter Values 163

4.3 Power-Law Index for Lognormal Function 163

4.4 Power Law Fit Values for Larson's Mass-Radius Relationship 175

Figures

Figure

1.1 Milky Way longitude-velocity diagrams from microwave emission lines. 6

2.1 Galactic mid-infrared emission model computed with Hyperion (Robitaille et al., 2012) viewed from the North Galactic Pole. 32

2.2 Foreground fraction of Galactic 8-μm emission in the northern Galactic plane derived from the Hyperion model as a function of (ℓ,$d_\odot$) along $b = 0°$. 33

2.3 BGPS and processed GLIMPSE data for example object G035.524-00.274. 35

2.4 Morphological matching for example object G035.524-00.274. 38

2.5 Longitude-velocity diagram of the northern Galactic plane. 41

2.6 Distributions of Galactic latitude, 40″ flux density, and mid-infrared contrast of sources. 45

2.7 Comparison of the full-width of the 68.3% error bar (FW_{68}) against the integrated DPDF on the d_{ML} side of $d_{\tan}$ (P_{ML}). 46

2.8 Comparison of source properties for objects with "near" vs. "far" KDA resolutions. 48

2.9 KDA resolution versus the ratio of BGPS 40″ flux density to mid-infrared contrast. . 49

2.10 Face-on view of the Milky Way for sources with well-constrained KDA resolutions, plotted atop an artist's rendering of the Milky Way (R. Hurt: NASA/JPL-Caltech/SSC) viewed from the north Galactic pole. 51

2.11 Vertical distribution of sources about the Galactic midplane. 51

2.12 Derived vertical position of sources versus heliocentric distance. 52

2.13 Comparison of KDA resolutions derived from the DPDF with published distance estimates. 54

2.14 Example DPDFs for three possible cases. 61

2.15 Histograms of EMAF contrast as a function of IRDC identification. 63

2.16 Azimuthally-averaged surface brightness ($\lambda = 8$ μm) as a function of galactocentric radius. 65

3.1 Galactic kinematics for distance determination. 87

3.2 Comparison of the $v_{\rm LSR}$ of the brightest ^{13}CO emission feature from the GRS data towards a BGPS object compared to the HCO$^+$(3-2) velocities of Shirley et al. (2013) and NH_3(1,1) velocities of Dunham et al. (2011b). 90

3.3 Demonstration of the morphological spectrum extraction technique for example source G035.523-00.273. 91

3.4 Flow chart for the assignment of $v_{\rm LSR}$ from the morphologically extracted ^{13}CO spectrum for a given molecular cloud structure. 93

3.5 Optimization of the $T_{A,\rm thresh}$ for detection and R_T limit for multiply peaked spectra. 94

3.6 Morphologically extracted ^{13}CO spectrum $v_{\rm LSR}$ comparison with dense-gas tracers. . 96

3.7 Optimization of the association volume for catalog-based priors, based on HRDS H II regions. 99

3.8 Example of DPDF$_{\rm px}$ construction from trigonometric parallax measurements. 101

3.9 Assignment of $v_{\rm LSR}$ as a function of $\lambda = 1.1$ mm peak flux density. 105

3.10 Longitude-velocity diagram for the BGPS. 108

3.11 The Gemini OB1 field. 113

3.12 "Well-constrained" distance estimate criteria. 114

3.13 Summary of source properties from Table 3.5. 119

3.14 Face-on view of the Milky Way for sources with well-constrained distance estimates, plotted atop an artist's rendering of the Milky Way (R. Hurt: NASA/JPL-Caltech/SSC) viewed from the north Galactic pole. 122

3.15 Vertical distribution of sources about the Galactic midplane. 122

3.16 Comparison of observable quantities for the entire BGPS catalog, the kinematic sample, and set of well-constrained distance estimates. 129

4.1 Comparison of mass distribution methods. 143

4.2 Computing mass completeness levels. 144

4.3 Brief analysis of resolved versus unresolved BGPS V2 sources. 151

4.4 Physical property histograms for the Distance Catalog. 152

4.5 Inter-comparison of the physical properties for sources from the BGPS Distance Catalog with finite θ_R. 154

4.6 Physical properties as a function of heliocentric distance and Galactocentric radius. . 156

4.7 Power-law fit to the aggregate mass distribution. 158

4.8 Lognormal fit to the aggregate mass distribution. 160

4.9 Simulated aggregate mass distributions of star-forming regions at a variety of distances.162

4.10 Astrophysically cut subsets. 165

4.11 Mass functions of subsets in heliocentric distance. 166

4.12 Mass functions of subsets in Galactocentric radius. 168

4.13 Mass functions of subsets in physical radius. 170

4.14 Mass functions of subsets in number density. 171

4.15 Mass functions of protocluster-type subsets. 173

4.16 The mass-size relationship for molecular cloud structures. 175

4.17 Dense gas mass surface density over the Galactic disk for Distance Catalog sources. . 177

4.18 Vertical and radial mass distributions. 180

4.19 Fraction of source type as a function of heliocentric distance. 182

4.20 Dense gas mass fraction estimate as a function of Galactocentric position. 190

Chapter 1

Introduction

1.1 Preface

To conclude, if we call **light**, those rays which illuminate objects, and **radiant heat**, those which heat bodies, it may be inquired, whether light be essentially different from radiant heat? In answer to which I would suggest, that we are not allowed, by the rules of philosophizing, to admit two different causes to explain certain effects, if they may be accounted for by one. (Herschel, 1800, p. 291)

So far as thought may peer into the past, the epic of our solar system began with a great catastrophe. Two suns met. What had been, ceased; what was to be, arose. Fatal to both progenitors, the event dated a stupendous cosmic birth. (Lowell, 1908, p. 3)

Not from the stars do I my judgment pluck;
And yet methinks I have astronomy,
But not to tell of good or evil luck,
Of plagues, of dearths, or seasons' quality;
Nor can I fortune to brief minutes tell,
Pointing to each his thunder, rain and wind,
Or say with princes if it shall go well,
By oft predict that I in heaven find:
But from thine eyes my knowledge I derive,
And, constant stars, in them I read such art
As truth and beauty shall together thrive,
If from thyself to store thou wouldst convert;
Or else of thee this I prognosticate:
Thy end is truth's and beauty's doom and date.
– **William Shakespeare, Sonnet 14**

For many years I have been a night watchman of the Milky Way galaxy. – **Bart Bok**

Sorry about the mess. – **Han Solo**

Galactic astronomy often takes a back seat to flashier topics such as black holes or cosmology. The Milky Way is a mundane spiral galaxy with a run-of-the-mill black hole at its center; no recent mergers, no massive starbursts, no active galactic nucleus. Boring, just like millions of other galaxies spread throughout the universe. But it is just this commonplaceness that makes it so interesting.

We live within a massive laboratory where nearly all of the physical processes relevant to the evolution of the universe are taking place at some level. There is a wealth of information surrounding us. The problem, however, is being able to see the forest through the trees.

This thesis makes one attempt at seeing through the trees. The process of star formation begins in giant molecular clouds and proceeds through feedback and gas expulsion to stars shining in the blackness. The fleeting intermediate steps have been so difficult to see or study, but enshroud the fundamental processes that make the universe as we see it today. Primarily due to the intersection of the cold temperatures of these regions and the only-recent availability of technology to fully survey the sky in the necessary portion of the electromagnetic spectrum, earnest study has only now begun.

Thanks to the wealth of information presently available, this thesis moves beyond the study of individual regions to develop a framework to systematically study the statistical properties of star-formation sites throughout the Milky Way, using the trees to build a picture of the forest.

1.2 Star Formation in the Milky Way

1.2.1 Overview of Star Formation

Stars drive the universe. Feedback from the first generations of stars precipitated the epoch of re-ionization and helped to define the structures we observe today (**cf.** Barkana & Loeb, 2001, and references therein). Despite stars (and in particular massive stars; $M \geq 8\,M_{\odot}$) being a major source of observable radiation in the universe and driver of structure, however, relatively little is known about the details of their formation.

In the most simplistic of views, giant molecular clouds (GMCs) form through converging flows (with multiple possible sources) and begin to fragment into Jeans-scale clumps. The clumps further fragment into self-gravitating cores that form individual stellar systems. The feedback from the new stars, in turn, expels the remaining molecular gas.

While this wide-angle picture is generally correct, the individual steps are shrouded in mystery. It is generally agreed that GMCs form through converging flows, but the nature of those flows may come from one or more of stellar feedback or turbulence, aggregation of smaller clouds, or (magneto-)gravitational and differential buoyancy instabilities (Dobbs et al., 2013). The clumpy, filamentary structure within GMCs may be caused by gravitational collapse (**e.g.**, Burkert & Hartmann, 2004), supersonic turbulent structure (**e.g.**, Padoan et al., 2001), internal colliding flows (**e.g.**, Vázquez-Semadeni et al., 2006), or multi-scale infall (**e.g.**, Schneider et al., 2010), where the clumpy structure is an observational artifact. Whatever the cause, the dense interstellar medium exhibits a fractal nature, wherein structures are self-similar and there appears to be no preferred scale between the bounds of the Galactic disk scale height and runaway gravitational collapse (Kennicutt & Evans, 2012).

One means for enumerating the distribution of mass amongst objects of various scales is the mass function, which describes the number of objects between M and $M + \delta M$. For GMCs, the mass function is described by a power law with $dN/dM \propto M^{-\alpha}$, with typical values of the power-law index $\alpha \approx 1.6 - 1.8$ (Blitz, 1993). For indices $\alpha < 2$, the majority of the mass resides in the largest few objects; most of the molecular mass in the Galaxy is contained in a handful of truly massive GMCs (**e.g.**, Sgr B2). At the other end of the star-formation engine, stars form with an apparently uniform initial mass function (IMF), which describes the relative numbers of massive versus low-mass stars (Offner et al., 2013). The IMF has a power-law tail to high mass with $\alpha = 2.35$ (Salpeter, 1955), indicating most of the stellar mass resides in the smallest stars. The progression from GMC mass function to the stellar IMF is unclear, but recent work has suggested that the mass function for **clumps** lies intermediate between these extremes (Donkov et al., 2012; Veltchev et al., 2013).

While there is still much uncertainty in the progression of massive star formation, it is generally agreed that most stars form in clusters (Lada & Lada, 2003) that correspond to the scale of molecular cloud clumps (Bergin & Tafalla, 2007). As the active, denser substructures within GMCs, these objects offer insight into the critical intermediate phase that sets the stage for runaway gravitational collapse and massive star formation (McKee & Ostriker, 2007). The difficulty, however, lies in systematically observing them within the Milky Way.

1.2.2 The Milky Way: Seeing the Forest Through the Trees

> One might have hoped that from this great wealth of information on the properties of the local ISM would have flowed a clear and definitive picture. Sadly this is not the case; studies of the Milky Way's ISM suffer badly from the syndrome of not being able to see the wood for the trees. The problem is two-fold. First, the wealth of the available information gets in the way of sweeping generalizations of the sort that may be plausibly made about external galaxies, about which we know so much less. Second, we have the misfortune to be located in the plane of the Milky Way, and suffer dreadfully from the particular bugbear of astronomy: ignorance of the distance to which we are peering. (Binney & Merrifield, 1998, p. 535)

Whereas for many external galaxies we have the luxury of seeing the face of an entire galactic disk with everything at the same distance, observations of the Milky Way greatly suffer from line-of-sight confusion and great ambiguity about the distances of observed objects. So, despite a multitude of surveys of the Galactic plane at many wavelengths observing the full range of Galactic ecology, there is still great uncertainty about the exact structure of the Galaxy. As noted in the quote above, distances to structures within the Galactic plane range from a few parsecs to tens of kiloparsecs. Continuum observations flatten this depth of structure onto the plane of the sky, and many creative methods have been developed over the decades to tease out information about the third dimension. Because stars obey an intrinsic temperature-luminosity relationship (**cf.** Russell, 1914), it is possible to determine distances by comparing stars' computed luminosity with the observed brightness. While this method requires modification due to interstellar reddening (dust extinction) at larger distances, it has greatly advanced the understanding of nearby Galactic structure.

The clouds of atomic and molecular gas that play host to star formation do not have intrinsic luminosity relationships that allow for distance estimation, so other techniques must be developed. The microwave transitions of atomic hydrogen (H I) and molecular species (**e.g.**, CO, H_2CO, etc.) offer a valuable observational third dimension that may be exploited: Doppler shifts. It was known from optical studies of external galaxies that spirals rotate with velocity extrema of hundreds of km s^{-1} (**cf.** Slipher, 1914). Extending this to the Milky Way, if the rotational velocity as a function of Galactocentric radius (the rotation curve) is measured or determined, then the Doppler shift of observed microwave transitions can be used to estimate distance from the Sun. Measuring line-of-sight, or radial, velocities ($v_{\rm LSR}$) with an accuracy of 5 km s^{-1} requires a spectral resolution $R = c/\Delta v = 6 \times 10^4$, or frequency resolution $\Delta\nu = \nu/R \approx 25$ kHz at the frequency of the ubiquitous H I 21-cm line, within the reach of technology available even 60 years ago (**cf.** Muller & Oort, 1951).

The primary difficulty in translating the observed $v_{\rm LSR}$ from these microwave emission lines into a heliocentric distance is an accurate rotation curve for the Milky Way. Distances estimated in this way are called "kinematic distances". Many rotation curves have been measured using various methods and data sets over the span of 60+ years (**e.g.**, Muller & Oort, 1951; Clemens, 1985; Brand & Blitz, 1993; Reid et al., 2014), and there remains uncertainty in the exact form and magnitude of the rotation. In all rotation curves, however, it is assumed that stars and gas rotate about the Galactic center in roughly circular orbits. In spite of this difficulty, much can be learned about Galactic structure in the absence of absolute distance information. Early spectroscopic surveys of the H I spin-flip transition ($\nu = 1420.4$ MHz) across the Galactic plane (**cf.** Ewen & Purcell, 1951) identified coherent structures in the longitude-velocity ($\ell - v$) diagram of the data indicating the presence of spiral structure (Schmidt, 1957). Later surveys of the molecular content of the Galaxy using ^{12}CO($J = 1 - 0$, $\nu = 115.27$ GHz) from the mid-1980s - 2000 (compiled in Dame et al., 2001) showed more-pronounced structure than the H I data (see Figure 1.1). Yet, even with such rich data in hand, the exact structure of the Galaxy is still unknown.

One of the chief difficulties in discerning Galactic structure in the inner Galaxy (**i.e.**, $|\ell| \leq 90°$)

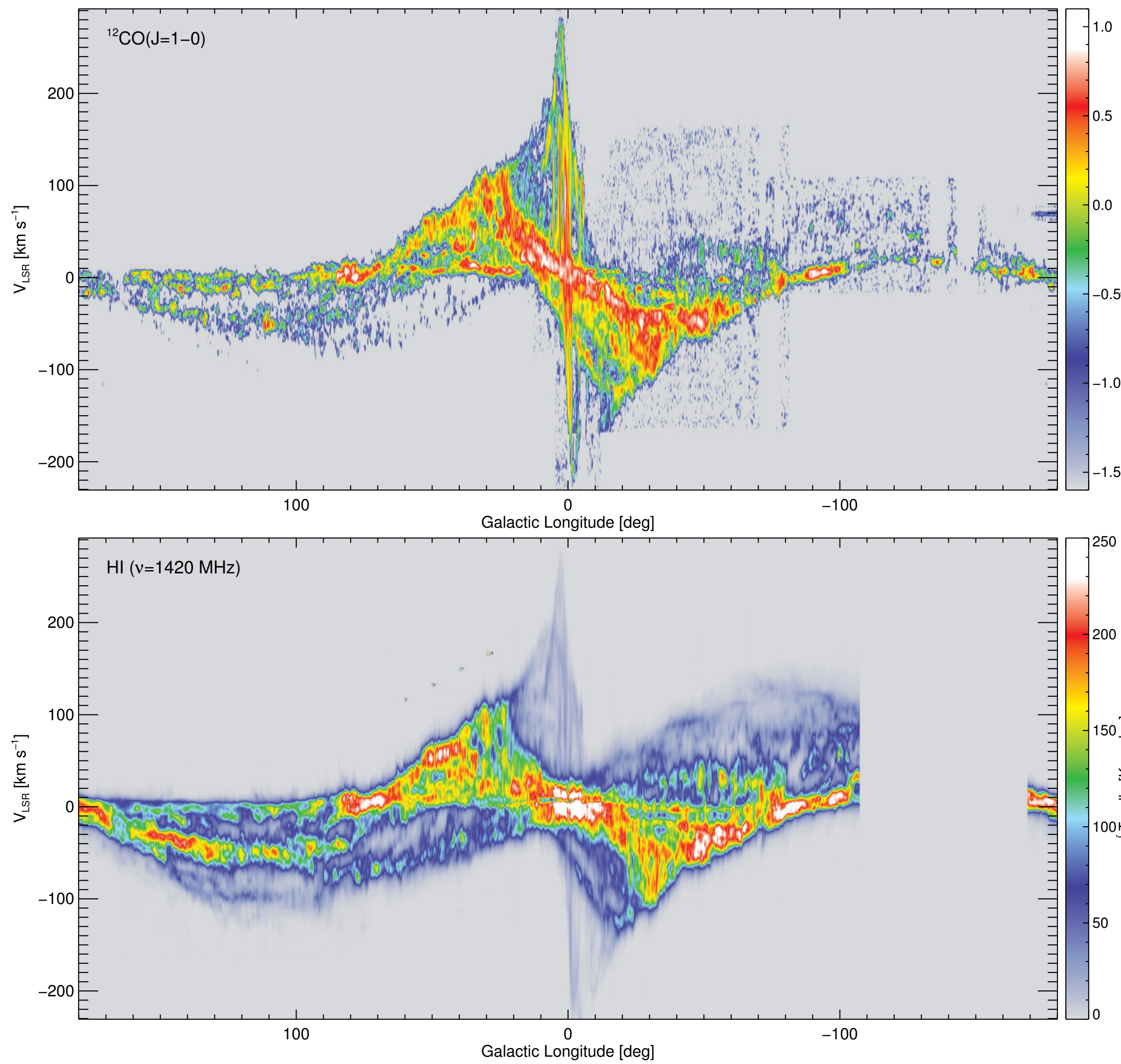

Figure 1.1 **Top**: Milky Way longitude-velocity diagram from the ^{12}CO(1-0) data of Dame et al. (2001). The full data cube has been integrated over the latitude range $-2^\circ \leq b \leq 2^\circ$ to produce the two-dimensional image. These data were downloaded directly from `http://www.cfa.harvard.edu/rtdc/CO/CompositeSurveys/`, and the moment-masking procedure suggested by Dame (2011) has not been applied. The lack of moment-masking leads to the "noise" in $\ell < 0^\circ$ regions outside the envelope of Galactic emission. This figure is styled after Dame et al. (2001, their Figure 3). **Bottom**: The longitude-velocity diagram derived from the H I ($\nu = 1420$ MHz) VLA Galactic Plane Survey (VGPS; Stil et al., 2006), Canadian Galactic Plane Survey (CGPS; Taylor et al., 2003), and Southern Galactic Plane Survey (SGPS; McClure-Griffiths et al., 2005). As with the CO data, the cubes have been integrated over the latitude range $-2^\circ \leq b \leq 2^\circ$ to produce the two-dimensional image. The gap at $250^\circ \geq \ell \geq 190^\circ$ is a lapse in coverage between these particular surveys; this longitude range is accessible only from the VLA, but no survey has been conducted to fill in this range. Note that the ^{12}CO data are shown on a logarithmic scale, but the H I data is on a linear scale.

from spectroscopic data arises from simple geometry. Regardless of the rotation curve chosen for converting $v_{\rm LSR}$ into position, the magnitude of $v_{\rm LSR}$ increases as a function of heliocentric distance until the point of closest approach between the line of sight and the Galactic center. At this point, the circular motion of gas and stars is tangent to the line of sight; this location is called the tangent point or tangent distance ($d_{\rm tan}$). Beyond $d_{\rm tan}$, the magnitude of $v_{\rm LSR}$ decreases, reaching 0 km s^{-1} upon meeting the solar circle, the circular orbit of the Sun about the Galactic center. Therefore, a single $v_{\rm LSR}$ may correspond to two distinct heliocentric distances, a difficulty known as the kinematic distance ambiguity (KDA).

As far back as Oort et al. (1958), it was recognized that the dominant drivers of spiral structure are not stars, but rather the gas and dust out of which the stars form. A comparison of the H I and CO $\ell - v$ diagrams in Figure 1.1 reveals a significant concentration of molecular gas into the spiral features, with little emission between. This is in contrast to the H I data, which indicate enhanced emission coincident with the prominent CO features, but is also more evenly spread through most of the Galactic plane. One notable exception is the near total lack of H I emission associated with the central molecular zone around $\ell = 0°$ where the CO emission spans a huge range of $v_{\rm LSR}$.

The coherent structures in Figure 1.1 may be fit with logarithmic spirals, yielding a means for converting the observed Milky Way into a face-on view analogous to external galaxies. Indeed, this sort of fitting was popularized by the work of Georgelin & Georgelin (1976), who supplemented the available H I data with subsequent surveys of Galactic H II regions and other tracers of Galactic structure. The fitting of logarithmic spirals to $\ell - v$ diagrams has proved useful for identifying new spiral arms in the extreme outer Galaxy (**cf.** Dame & Thaddeus, 2011) and making arguments about the existence of the apparent Molecular Ring, an enhancement of molecular gas at a Galactocentric radius of 4-5 kpc (**cf.** Dobbs & Burkert, 2012).

1.3 Observations of Star Formation Tracers

1.3.1 How Do We Observe Star Forming Regions?

Giant molecular clouds (GMCs), the hosts of star formation, are typically cold ($T \approx 10-40$ K) structures. As such, blackbody emission from dust grains embedded therein peaks at sub-millimeter wavelengths and is not observable far from this bandpass. Furthermore, thermal conditions are generally not sufficient to excite the molecular gas above the lowest rotational quantum states across large swaths of the Galactic disk. Emission from the lowest rotational transitions of the simplest (and therefore most abundant) molecules typically have $\lambda \sim 1$ mm, plus or minus an order of magnitude. GMCs may be studied via either method, both of which have advantages and difficulties. Dust continuum emission at these wavelengths is optically thin, and emission may be detected from sources across the Galactic disk. In particular, there is no critical density at which thermal dust emission turns on, and only in the most extreme of environments (**e.g.**, the Sgr B2 molecular cloud; Bally et al., 2010b) does high density threaten the optically-thin assumption. Continuum surveys, however, are only sensitive to plane-of-sky location of emission; information about the third dimension must be gleaned from ancillary data. Molecular transition lines, on the other hand, have the Doppler shift v_{LSR} imprinted in the frequency spectrum, providing a foothold on distance estimation. The downside comes as the transitions of different molecular species are sensitive to different density regimes within GMCs, and those with low critical density for emission (**e.g.**, CO(J=1-0)) rapidly become optically thick. Therefore, a suite of surveys, observing molecular transition lines sensitive to a broad range of density conditions, are required to encompass the required dynamic range.

This microwave / terahertz section of the electromagnetic spectrum has been one of the last to be explored astronomically due to the technical difficulties of detecting the light by either continuum bolometers or heterodyne receivers. Early bolometers were typically single-pixel instruments and were used to study individual known regions of cold molecular gas. Likewise heterodyne receivers that were able to work above 100 GHz were noisy instruments with limited sensitivity. Both types of instruments would have required impractical amounts of observing time to complete a survey of the Galactic plane.

1.3.2 The Bolocam Galactic Plane Survey

The advent of large-format bolometer arrays in the early 2000's greatly increased observing efficiency and allowed for the possibility of conducting wide-scale blind surveys of the sky. While several have now been completed or are in progress (**e.g.**, ATLASGAL, Schuller et al., 2009; Hi-GAL, Molinari et al., 2010b), the Bolocam Galactic Plane Survey (BGPS; Aguirre et al., 2011; Ginsburg et al., 2013) was the first to be completed. The BGPS is a $\lambda = 1.1$ mm continuum survey that utilized the Bolocam[1] instrument (a bolometer array instrument with a filter designed to exclude the bright ^{12}CO(J=2-1) emission line at 230.5 GHz; Glenn et al., 2003) at the Caltech Submillimeter Observatory (CSO) on Mauna Kea. Data were collected between 2005 July and 2007 September, with additional fields observed in 2009 December. Although Bolocam has an angular resolution of $31''$, optical distortions unaccounted for in the pipeline degrade the effective resolution of the BGPS mosaics to $33''$.

The BGPS consists of a contiguous blind survey of the first quadrant of the Galactic plane from $-10° \leq \ell \leq 90°$ covering $|b| \leq 0\overset{\circ}{.}5$, with flares out to $|b| \leq 1\overset{\circ}{.}5$ at several longitudes. In addition, a series of targeted regions in the outer Galaxy (**e.g.**, IC1396, W3/4/5, Gemini OB1) were observed individually, chosen based on CO surveys. Version 1 (V1) of the data, covering ≈ 170 deg^2, was released in 2009 and is described, along with the custom data-reduction pipeline in Aguirre et al. (2011). When compared to other published data at comparable wavelengths, however, the V1 data had measured flux densities that were too low by a factor of 1.5, prompting the recommendation of a correction factor to be applied to the map data and cataloged flux densities. This flux density deficiency led to a reworking of the BGPS pipeline and subsequent re-reduction of the data, released in 2013 as version 2 (V2; Ginsburg et al., 2013), where an incorrect calibration value was identified as the source for the deficient flux densities in the V1 data. In addition to an

[1] `http://www.cso.caltech.edu/bolocam`

improved data reduction, the V2 release added several more fields in the outer Galaxy, bringing the total observed area to ≈ 192 deg^2.

A custom cataloging routine was written to extract sources from BGPS surveys mosaics, and details may be found in Rosolowsky et al. (2010b). The cataloger is based on a seeded-watershed algorithm, which identifies regions of significant emission in a map and assigns each pixel that is likely to contain emission to a single object based on the structure of the emission and nearby local maxima. The V1 catalog contains 8,358 entries, but was superseded by the V2 catalog, which contains 8,594 sources extracted from the V2 maps. While the primary differences between the two catalogs come from the expanded coverage in V2, the improved noise characteristics and flux-density recovery of the newer maps led to changes in source identification (see Ginsburg et al., 2013 for complete details). For instance, some monolithic V1 sources were subdivided into multiple V2 sources or multiple V1 sources combined into a single V2 source because of the improved measurement of emission topography. Additionally, new sources were identified at lower signal-to-noise in the V2 images due to improved noise characteristics, and apparently spurious V1 sources were omitted that did not meet the signal-to-noise thresholds.[2]

The third dimension has been added to the BGPS data through several follow-up spectroscopic surveys to derive v_{LSR} for the structures identified in the source catalog. Since the plane-of-sky locations for dense regions in GMCs were identified by the continuum data, spectroscopic follow-up observations were pointed only at V1 (or earlier) catalog objects, greatly speeding the process. Velocity catalogs have been published for NH_3(1,1) (near 24 GHz; Dunham et al., 2010, 2011b) and HCO^+(3-2) and N_2H^+(3-2) (near 270 GHz; Schlingman et al., 2011; Shirley et al., 2013). Additional observations have been conducted using CS(2-1) (near 100 GHz; Y. Shirley, 2012, private communication) and $C^{18}O$(2-1) (near 220 GHz; M. Lichtenberger, 2014, private communication). Along with a novel method for extracting a pseudo-dense-gas spectrum from ^{13}CO (presented in

[2] Shirley et al. (2013) identify several V1 sources with positive detection it the dense-gas tracer HCO^+(3-2) that are absent from the V2 catalog. This suggests that omitted V1 sources are not entirely spurious, and that the cataloging routine may benefit from additional tuning to best match the characteristics of the V2 images. Furthermore, future improvements to the cataloging routine may better distinguish between compact and filamentary sources and produce a catalog with higher fidelity to the observed structure.

Chapter 3), these velocity catalogs provide the basis for computing kinematic distances for nearly 50% of the V2 catalog.

Since the BGPS and other similar surveys are cataloging for the first time the dense molecular cloud structures that will give rise to collections of stars, capitalization on the data set is crucial. Distance estimates are critical for deriving the physics of these objects, and several avenues have been pursued for utilization of the BGPS data. The first is to find a collection of objects at a known distance, such as the Galactic center and central molecular zone (Bally et al., 2010b) or the Gemini OB1 molecular cloud (Dunham et al., 2010). Second, kinematic distances may be used if the KDA may be resolved, and a first attempt was made using NH_3 velocities by Dunham et al. (2011b) and with HCO^+(3-2) and N_2H^+(3-2) velocities by Schlingman et al. (2011). Both of these studies relied in part on association with infrared dark clouds (IRDCs), which are assumed to be at the near kinematic distance (see Chapter 2 for a discussion of this assumption and its limitations), and Dunham et al. additionally used a technique of searching for absorption features in H I 21-cm spectra.

The physical properties of objects are likely influenced by their environments and the degree of star-formation activity taking place within. Dunham et al. (2011a) compiled a census of active star-formation tracers for the BGPS V1 catalog sources, dividing sources into groups from active to no observable signs. Additionally, since the BGPS angular resolution is somewhat coarse (0.8 pc at a distance of 5 kpc), many of the cataloged sources are composed of multiple smaller structures (Merello et al., 2014). Ultimately, observations with ALMA will quantify the continuum of structure and scaling relationships for dense structures within molecular clouds and the importance and relevance of various scales to the process of star formation.

1.4 Scope of Work

1.4.1 A Probabilistic Distance Estimation Method

My arrival to the BGPS collaboration in 2010 coincided with the public release of the images and catalog and the first round of publications describing the data and early results. I was given the task of bringing to maturation a novel concept for probabilistic distance estimation that could be applied to the entire catalog. Later given the name distance probability density functions (DPDFs), the method involves combining as many sources of information relevant to the distance as possible in a Bayesian framework (Bayes & Price, 1763). The fundamental equation for this work is

$$\mathrm{DPDF}(d_\odot) = \mathcal{L}(v_{\mathrm{LSR}}, l, b; d_\odot) \prod_i P_i(l, b; d_\odot) \ , \tag{1.1}$$

where $\mathcal{L}(v_{\mathrm{LSR}}, l, b; d_\odot)$ is the kinematic distance likelihood function 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 details of using Equation (1.1) are described in Chapters 2 and 3.

The work for this thesis began with combining several small extant routines to compute kinematic distances and simple priors into a coherent set of code that could run nearly autonomously on large catalogs of objects. After many iterations and revisions, that code will be released to the public[3] on publication of Ellsworth-Bowers et al. (2014). The workings and features of the code are described in Appendix A.

The two terms in Equation (1.1) both have a non-trivial amount of work behind them. At the start of my work with DPDFs, less than 2,000 BGPS V1 sources had been observed in one of the spectroscopic follow-up programs (Dunham et al., 2010, 2011b; Schlingman et al., 2011). To compute a maximally useful set of distance estimates, larger spectroscopic campaigns were needed to compute kinematic distance likelihood functions. I therefore participated heavily in the complete HCO^+ (3-2) / N_2H^+ (3-2) survey of all BGPS V1 sources at $\ell \geq 7\overset{\circ}{.}5$ published in Shirley et al. (2013),[4] where the velocity detection rate for the full survey was $\approx 50\%$. To further expand the number of

[3] `https://github.com/BGPS/distance-omnibus`

[4] The longitude limit was the lowest reliably observable from the HHT on Mt. Graham, Arizona.

objects with a measured $v_{\rm LSR}$, I refined and optimized a method (introduced in Rosolowsky et al., 2010a) for extracting the spectrum from dense structures within molecular clouds from ^{13}CO(1-0) data (Ellsworth-Bowers et al., 2014). The combination of velocity information from the various dense-gas surveys and ^{13}CO(1-0) was ultimately applied to the BGPS V2 catalog (see Chapter 3).

The other term in Equation (1.1) is of critical importance for objects in the inner Galaxy ($|\ell| \leq 90°$). Because of the KDA, $\mathcal{L}(v_{\rm LSR}, l, b; d_\odot)$ are double-peaked for these objects, and additional information is required to produce a single-peaked posterior DPDF. I developed two new powerful prior DPDFs and included a prior based on the distribution of molecular gas in the Galactic disk introduced in Rosolowsky et al. (2010a). The first new prior is an extension of the IRDC methodology employed by Dunham et al. (2011b) and Schlingman et al. (2011) using mid-infrared extinction as a means for resolving the KDA. Whereas previous studies utilized published catalogs of IRDCs, this new prior begins with dense molecular cloud structures identified at millimeter wavelengths and searches for an (eight-micron) absorption feature (EMAF) against the ubiquitous Galactic emission from polycyclic aromatic hydrocarbons (PAHs) near $\lambda = 8\ \mu$m (Ellsworth-Bowers et al., 2013). The second prior is a means for associating robust distance estimates from the literature (including trigonometric parallax measurements of masers in star formation regions) with dense molecular cloud structures (Ellsworth-Bowers et al., 2014), greatly expanding the number of BGPS objects with well-constrained distances.

1.4.2 Galactic Structure and the Physical Properties of Dense Molecular Cloud Structures

The real payoff for measuring distances to dense molecular cloud structures is derivation of mass and other physical properties. The original goal of this work was to measure the mass function of molecular cloud clumps and compare that to the mass function of GMCs and the stellar initial mass function (IMF). As work progressed, it became apparent that another significant payoff is the derivation of the Galactic distribution of dense molecular gas and comparisons with other tracers of Galactic structure. Furthermore, physical properties, such as the physical radius and number density of detected objects, can further investigate scaling relationships, such as those originally discussed by Larson (1981). This work is nearing fruition (T. Ellsworth-Bowers, in preparation); the preliminary version of the results and discussion are presented in this thesis.

1.5 Organization

This thesis consists of three papers prepared for the Astrophysical Journal. While the author lists on these papers are extensive, material assistance came in the form of ideas, small pieces of code, and comments on the written papers; the work contained in this thesis is largely the labor of a single person. Chapter 2 introduces the DPDF methodology and initial priors based on a model of the distribution of molecular gas in the Galactic disk and on morphological matching between millimeter continuum emission and EMAFs identified in the **Spitzer**/GLIMPSE survey images (Ellsworth-Bowers et al., 2013). Chapter 3 expands the DPDF formalism to include a kinematic distance likelihood from ^{13}CO(1-0) data and the prior based on robust distance estimates from the literature to present a distance catalog for 1,802 BGPS V2 sources across the Galactic plane (Ellsworth-Bowers et al., 2014). Chapter 4 uses the distance catalog to compute the physical properties of dense molecular cloud structures and investigate the mass function(s) of these objects (T. Ellsworth-Bowers, in preparation). In all three chapters, Galactic structure is discussed as relevant to the content of the chapter. Finally, Chapter 5 provides a summary of the overarching results and themes from this work and provides a brief outline of future work to be done in applying the methodology presented here to the entire Galactic plane through other (sub-)millimeter surveys and expanding the suite of available prior DPDFs to produce well-constrained distances for a larger fraction of source catalogs. Appendix A describes the `distance-omnibus` code base I developed for this work and will release to the public.

Chapter 2

A Mid-Infrared Kinematic Distance Discrimination Method

We present a new distance estimation method for dust-continuum-identified molecular cloud clumps. Recent (sub-)millimeter Galactic plane surveys have cataloged tens of thousands of these objects, plausible precursors to stellar clusters, but detailed study of their physical properties requires robust distance determinations. We derive Bayesian distance probability density functions (DPDFs) for 770 objects from the Bolocam Galactic Plane Survey in the Galactic longitude range $7\overset{\circ}{.}5 \leq \ell \leq 65°$. The DPDF formalism is based on kinematic distances, and uses any number of external data sets to place prior distance probabilities to resolve the kinematic distance ambiguity (KDA) for objects in the inner Galaxy. We present here priors related to the mid-infrared absorption of dust in dense molecular regions and the distribution of molecular gas in the Galactic disk. By assuming a numerical model of Galactic mid-infrared emission and simple radiative transfer, we match the morphology of (sub-)millimeter thermal dust emission with mid-infrared absorption to compute a prior DPDF for distance discrimination. Selecting objects first from (sub-)millimeter source catalogs avoids a bias towards the darkest infrared dark clouds (IRDCs) and extends the range of heliocentric distance probed by mid-infrared extinction and includes lower-contrast sources. We derive well-constrained KDA resolutions for 618 molecular cloud clumps, with approximately 15% placed at or beyond the tangent distance. Objects with mid-infrared contrast sufficient to be cataloged as IRDCs are generally placed at the near kinematic distance. Distance comparisons with Galactic Ring Survey KDA resolutions yield a 92% agreement. A face-on view of the Milky Way using resolved distances reveals sections of the Sagittarius and Scutum-Centaurus Arms. This

KDA-resolution method for large catalogs of sources through the combination of (sub-)millimeter and mid-infrared observations of molecular cloud clumps is generally applicable to other dust-continuum Galactic plane surveys.

2.1 Introduction

Recent (sub-)millimeter surveys of the Galactic plane (ATLASGAL, Schuller et al., 2009; Hi-GAL, Molinari et al., 2010b; BGPS, Aguirre et al., 2011) have detected tens of thousands of molecular cloud cores and clumps in thermal dust emission. As plausible precursors to stellar clusters, OB associations, or smaller stellar groups, molecular cloud clumps can yield clues about the formation of massive stars (McKee & Ostriker, 2007). The masses and temperature profiles of these objects are key to unraveling this process. Recent work has sought to measure these quantities (Russeil et al., 2011; Eden et al., 2012), but a robust and comprehensive tally does not yet exist.

Derivation of masses for molecular cloud clumps from dust continuum data requires an estimate of the heliocentric distance to each object and the temperature of the emitting dust. Analysis of **Herschel** Hi-GAL data is beginning to yield temperature maps of the Galactic plane (Peretto et al., 2010; Battersby et al., 2011). While a detailed understanding of the interplay between dust temperature and the environment and evolution of molecular cloud clumps is important, variations in the assumed dust temperature by a factor of two only produce a factor of a few difference in the mass derived from (sub-)millimeter observations. In contrast, the derived mass of a molecular cloud clump is proportional to the square of its heliocentric distance; accurate distance estimates play a far larger role in the mass calculation. Recent studies of isolated regions, with well-determined distances, such as Perseus and Ophiuchus (Ridge et al., 2006; Enoch et al., 2006; Rosolowsky et al., 2008a), have unveiled many properties of molecular cloud cores in recent years (Enoch et al., 2007; Schnee et al., 2010). To gain similar insight into the larger molecular cloud clumps seen spread throughout the Galactic plane, a robust method for distance determinations for large data sets is required, because the distances to most clumps are subject to the kinematic distance ambiguity (KDA).

The most straightforward method for estimating the heliocentric distance ($d_\odot$) to a molecular cloud clump is to project its observed line-of-sight velocity ($v_{\rm LSR}$), derived from molecular line Doppler shifts, onto a Galactic rotation curve. These kinematic distances are generally unique for the outer Galaxy, but inner Galaxy sources are subject to the KDA, a projection effect of the orbital motion for objects within the Solar Circle (R_0). A line of sight intersecting a circular orbit at Galactocentric radius $R_{\rm gal} < R_0$ crosses that orbit twice, each with different spatial velocities but both with the same $v_{\rm LSR}$. Various techniques have been suggested for resolving the KDA (21-cm H I absorption: Anderson & Bania, 2009, Roman-Duval et al., 2009; the presence of mid-infrared dark clouds: Rathborne et al., 2006, Peretto & Fuller, 2009; H_2CO absorption: Sewilo et al., 2004; and near-infrared extinction: Marshall et al., 2009, Foster et al., 2012); this paper presents a method based on comparing mid-infrared extinction with (sub-)millimeter emission.

Appearing as dark absorption features against a bright mid-infrared background, infrared dark clouds (IRDCs) offer a practicable means for resolving the KDA. IRDCs are most striking against the broad, diffuse Galactic emission near $\lambda = 8$ μm (Perault et al., 1996; Simon et al., 2006a), although they may be detected in absorption against background stars at other infrared wavelengths (**cf.** Foster et al., 2012). Studies of IRDCs at (sub-)millimeter wavelengths reveal that they are dense molecular cloud clumps (Johnstone et al., 2003; Rathborne et al., 2006; Battersby et al., 2010, 2011). As extinction features, IRDCs must lie in front of enough mid-infrared emission to be visible. It is possible to **a priori** assign the near kinematic distance for the darkest clouds (**e.g.**, Butler & Tan, 2009; Peretto & Fuller, 2009), but recent work by Battersby et al. (2011) has shown that molecular cloud clumps may be visible as slight intensity decrements in the mid-infrared at the far kinematic distance despite not being dark enough to be cataloged as IRDCs. To encompass this second set of objects, we classify all dust-continuum-identified molecular cloud clumps with mid-infrared intensity decrements of any amount as Eight-Micron Absorption Features (EMAFs), whether catalogued as either an IRDC or not. These constitute a generalized collection of cold molecular cloud clumps identified first by dust-continuum emission and then checked for infrared absorption. The EMAF definition excludes objects extensively undergoing the later stages of star formation or that are exposed to strong ultraviolet radiation, as both processes excite PAH emission near $\lambda = 8$ μm, rendering invisible any absorption. Investigating the mid-infrared properties of molecular cloud clumps based on this classification avoids a bias toward the darkest, nearby IRDCs.

This paper presents a quantitative distance estimation technique for molecular cloud clumps based on Bayes' Theorem. A distance probability density function (DPDF) is computed using a distance likelihood derived from kinematic information (observed $v_{\rm LSR}$) and prior probabilities, based on ancillary data sets, that are applied in an effort to resolve the KDA. We present here two such priors. The first involves the comparison between observed mid-infrared absorption and millimeter emission of individual molecular cloud clumps, and the second is based on the Galactic-scale distribution of molecular gas. In addition to those described here, any number of additional priors may be applied to constrain the distance estimate.

The apparent optical depth of an EMAF calculated naïvely from mid-infrared images is likely less than the true value due to diffuse 8-μm emission lying between the cloud and the observer. By parameterizing the amount of total mid-infrared emission along a line of sight lying in front of a molecular cloud clump as the "foreground fraction" ($f_{\rm fore}$), simple radiative transfer arguments may be used to derive the true optical depth. The recent numerical Galactic infrared emission model of Robitaille et al. (2012) offers an estimate of $f_{\rm fore}$ as a function of $d_\odot$ in the Galactic plane. The maximum likelihood distance to a molecular cloud clump may be derived by comparing the optical depth calculated from (sub-)millimeter thermal dust continuum data with the absorption optical depth derived from the mid-infrared images and $f_{\rm fore}(d_\odot)$. This comparison generates a DPDF that takes into account Galactic-scale conditions along a given line of sight, including spiral structure. A DPDF derived in this manner contrasts the widely-used "step-function" method whereby a molecular cloud clump is automatically assigned the near kinematic distance upon association with a catalogued IRDC.

The methodology presented here is valid only for molecular cloud clumps that exhibit mid-infrared absorption, and therefore is but one means for distance discrimination for large catalogs of

Table 2.1. Spectroscopic Follow-up Observations of Inner Galaxy BGPS Sources

Species	Transition	ν (GHz)	Resol.[a] ($''$)	n_{eff}[b] (cm^{-3})	$\mathrm{N_{BGPS}}$[c]	Ref.
HCO^+	$J = 3 - 2$	267.6	28	10^4	6194	1
N_2H^+	$J = 3 - 2$	279.5	27	10^4	6194	1
CS	$J = 2 - 1$	97.98	64	5×10^3	553	2
NH_3	(1,1)	23.69	31	10^3	631	3

[a]Beam FWHM

[b]Approximate effective density for line excitation at $T = 20$ K (Evans, 1999)

[c]Number of unique BGPS sources observed in this line

References. — (1) Shirley et al. (2013); (2) Y. Shirley (2012, private communication); (3) Dunham et al. (2011b)

dust-continuum-identified objects. We present an automated means for deriving Bayesian DPDFs for mid-infrared dark molecular cloud clumps detected by the Bolocam Galactic Plane Survey, but this method is applicable to all (sub-)millimeter Galactic plane surveys.

This chapter is organized as follows. Section 2.2 describes the data sets used. The DPDF formalism is described in Section 2.3. Section 2.4 outlines the generation of prior DPDFs for EMAFs. Results from the Bayesian DPDFs are presented in Section 2.5. Implications of this work are discussed in Section 2.6, and a summary is presented in Section 2.7.

2.2 Data Sets

2.2.1 The Bolocam Galactic Plane Survey

The Bolocam Galactic Plane Survey (BGPS; Aguirre et al., 2011; Ginsburg et al., 2013) is a $\lambda = 1.1$ mm continuum survey covering 170 deg^2 at $33''$ resolution. The BGPS was observed with the Bolocam instrument at the Caltech Submillimeter Observatory (CSO) on Mauna Kea. 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 $|b| \leq 0\overset{\circ}{.}5$, plus selected regions in the outer Galaxy. For a map of

BGPS V1.0 coverage and details about observation methods and the data reduction pipeline, see Aguirre et al. (2011, hereafter A11).

From the BGPS V1.0 images, 8,358 millimeter dust-continuum sources were identified using a custom extraction pipeline. The BGPS catalog (Bolocat) contains source positions, sizes, and flux densities extracted in various apertures, among other quantities (see Rosolowsky et al., 2010b for complete details). BGPS V1.0 pipeline products, including image mosaics and the catalog, are publicly available[1] . For this work, we utilized the flux densities measured in a $40''$ top-hat aperture, which has the same solid angle as the BGPS $33''$ FWHM Gaussian beam ($\Omega = 2.9 \times 10^{-8}$ sr), in addition to the map data. A flux calibration multiplier of 1.5 ± 0.15 was applied to both Bolocat and the image mosaics to correct a V1.0 pipeline error (see A11 and Ginsburg et al., 2013 for a full discussion).

The BGPS data pipeline removes atmospheric signal using a principle component analysis technique that discards time-stream signals correlated spatially across the bolometer array. This effectively acts as an angular filter, attenuating angular scales comparable to or larger than the array field of view (see A11, their Fig. 15). The implication is that the BGPS is not sensitive to scales larger than $6'$. The effective angular size range of detected BGPS sources therefore corresponds to anything from molecular cloud cores up to entire clouds depending on the heliocentric distance (Dunham et al., 2011b). In this work we refer to BGPS objects as "molecular cloud clumps" for simplicity, but recognize that distant sources are likely larger structures.

2.2.2 Spectroscopic Follow-Up of BGPS Sources

Several spectroscopic follow-up programs have been conducted to observe BGPS sources in a variety of molecular emission lines that trace the dense gas associated with molecular cloud clumps. These surveys provide both kinematic and chemical information, and are typically beam-matched to the BGPS to facilitate comparison to the dust-continuum data. From these observations, a

[1] Available through IPAC at
`http://irsa.ipac.caltech.edu/data/BOLOCAM_GPS`

line-of-sight velocity ($v_{\rm LSR}$) was successfully fitted for each of more than 3,500 detected sources. A summary of spectroscopic programs is presented in Table 2.1.

In a pilot study (Schlingman et al., 2011) and complete survey (Shirley et al., 2013), all 6,194 Bolocat objects at $\ell \geq 7\overset{\circ}{.}5$ were observed using the Heinrich Hertz Submillimeter Telescope (HHT) on Mt. Graham, Arizona. These studies simultaneously observed the J=3−2 rotational transitions of HCO^+ ($\nu = 267.6$ GHz) and N_2H^+ ($\nu = 279.5$ GHz). Because these molecular transitions trace fairly dense gas ($n_{\rm eff} \approx 10^4$ cm^{-3})[2] , the line-of-sight confusion seen in CO studies is largely absent. In fact, Shirley et al. find only 2.5% of HCO^+ detections have multiple velocity components. These objects, likely an overlap of two or more molecular cloud clumps along the line of sight, are not used in this study. Detectability in HCO^+ is a strong function of millimeter flux density, and the detection rate for the full HHT survey was $\approx 50\%$ (see Shirley et al., 2013 for full details). Velocity fits to HCO^+ spectra constitute the bulk of the kinematic data used in this study (N_2H^+ spectra were not used because the complex hyperfine structure of its transitions makes it difficult to fit $v_{\rm LSR}$).

As a companion to the HHT observations, a subset of 555 BGPS sources were observed in the J=2−1 rotational transition of CS ($\nu = 97.98$ GHz) using the Arizona Radio Observatory 12m telescope on Kitt Peak (Y. Shirley 2012, private communication; see Bally et al., 2010a). This subset was confined to $29^\circ \leq \ell \leq 31^\circ$, a region with a high density of sources looking toward the Molecular Ring and the end of the long Galactic bar. This transition of CS traces lower density gas ($n_{\rm eff} \approx 5 \times 10^3$ cm^{-3}) than $HCO^+(3-2)$, and was detected in 45% of sources not detected by the HHT survey in this region.

Seeking to characterize the physical properties of BGPS sources, Dunham et al. (2011b) used the Robert F. Byrd Green Bank Telescope to observe the lowest inversion transition lines of NH_3 near 24 GHz. They observed 631 BGPS sources in the inner Galaxy. The NH_3 (1,1) inversion is the strongest ammonia transition at the cold temperatures of BGPS sources ($T \approx 20$ K), and we

[2] The effective density required to produce line emission with a brightness temperature of 1 K; may be up to several orders of magnitude smaller than the critical density (Evans, 1999).

used this transition exclusively for the NH_3 velocity fits.

2.2.3 The Spitzer GLIMPSE Survey

The **Spitzer** GLIMPSE survey (Benjamin et al., 2003; Churchwell et al., 2009) was used to identify mid-infrared extinction features associated with BGPS detected sources. The GLIMPSE survey area completely encompasses the BGPS for $|b| \leq 1\overset{\circ}{.}0$ and $\ell \leq 65°$(there are several sections of the BGPS that flare out to $|b| \leq 1\overset{\circ}{.}5$, see A11). We used the V3.5 IRAC Band 4 mosaics[3] ($\lambda_c = 7.9$ μm) to identify absorption features. Point sources (stars) identified in the Band 1 mosaics ($\lambda_c = 3.6$ μm) were removed from the Band 4 images to accentuate diffuse emission (see §2.4.2). Stars were modeled as Gaussian peaks since the mosaicing process from individual IRAC frames produces a spatially variable PSF, hampering star-subtraction. The Band 4 mosaics have an angular resolution $\sim 2''$, and a pixel scale of $1\overset{''}{.}2$. GLIMPSE images have undergone zodiacal light subtraction based on a zodiacal emission model (see the data product manual[3]), so signal remaining in the mosaics is Galactic in nature. There is, however, a significant effect due to scattering of light within the IRAC camera that causes the surface brightness of extended emission to appear brighter than it actually is[4] (Reach et al., 2005). The method used in this study to correct for scattered light is described in §2.4.2, and a derivation of the correction factors required for quantities measured from the publicly-available GLIMPSE mosaics is given in Appendix 2.8.2.

2.3 Distance Probability Density Functions

2.3.1 Approach and Utility

We introduce an automated distance determination technique for molecular cloud clumps that allows for the joint application of many individual distance estimation methods. Bayes' Theorem provides a framework for creating distance probability density functions (DPDFs) for

[3] Data product manual:
`http://irsa.ipac.caltech.edu/data/SPITZER/GLIMPSE/doc/glimpse1_dataprod_v2.0.pdf`

[4] See §4.11 of `http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/iracinstrumenthandbook/`

dust-continuum-identified molecular cloud clumps that encode the confidence in source distances. Kinematic distances derived from $v_{\rm LSR}$ and a Galactic rotation curve constitute the likelihood functions in the Bayesian context. Because these likelihoods are subject to the KDA, prior DPDFs based on ancillary data must be applied to constrain the distance estimates. The posterior DPDF is simply the product of the likelihood with the priors, suitably normalized. Relative amplitudes of the posterior DPDF at each distance along the line of sight ($d_\odot$) correspond to the probability of the source being at that distance.

Within this framework, any number of prior DPDFs may be applied to constrain the distances to molecular cloud clumps. This paper describes two such priors. The first, applicable to all molecular cloud clumps, is based on the Galactic distribution of molecular hydrogen. Because the scale height of the molecular disk is small, this prior favors the near kinematic distances for objects at high Galactic latitudes. The second prior involves the use of EMAFs. Not all molecular cloud clumps are visible as absorption features, however, so this prior (described in detail in §2.4) applies only to a subset of objects. To expand the collection of molecular cloud clumps with well-constrained DPDFs, additional techniques (**e.g.**, HISA, NIREX, etc.) would need to be applied.

Not only do DPDFs provide a structure for applying multiple techniques for distance discrimination, they also encode the distance uncertainty and level of confidence in the KDA resolution. When used to derive the mass or other property of a molecular cloud clump, DPDFs provide a means for determining the associated uncertainty. The DPDFs derived in this work are computed out to a heliocentric distance of 20 kpc in 20-pc intervals. To facilitate the use of integrated probabilities, DPDFs are normalized to unit total probability such that $\int_0^\infty \mathrm{DPDF}\ \mathrm{d}(d_\odot) = 1$.

2.3.2 Extracting a Distance from the DPDF

The proper use of DPDFs for calculating derived quantities is to build a distribution by randomly sampling distances from the DPDFs in a Monte Carlo fashion, preserving all information about distance placement and uncertainty. There are applications, however, that benefit from or require a single distance estimate with uncertainty (such as distance comparisons with other studies). There are two primary distance estimates that may be derived from a DPDF. The maximum-likelihood distance (d_{ML}) is the distance which maximizes the DPDF. This represents the single best-guess at the distance for cases where a large fraction of the total probability lies within a single peak. The associated uncertainty may be defined as the confidence region around d_{ML} that encloses at least 68.3% of the integrated DPDF, and whose limits occur at equal relative probability. This so-called isoprobability confidence region is generally asymmetric, and may represent lopsided error bars several kiloparsecs in size if both kinematic distance peaks are required to enclose sufficient probability. The full width of this uncertainty (FW_{68}), therefore, provides a direct measure of how well constrained a distance estimate is. Error bars produced in this way should not be considered Gaussian, as the 95.5% and 99.7% isoprobability confidence regions may be similar in size to the 68.3% error bars, or be radically different.

An alternative single-value distance estimate is the weighted average distance ($\overline{d}$), the first moment of the distribution,

$$\bar{d} = \int_0^\infty d_\odot \ \mathrm{DPDF} \ \mathrm{d}(d_\odot) \ . \tag{2.1}$$

If the DPDF is well-constrained to a single peak, d_{ML} and $\overline{d}$ will be nearly equivalent. In cases where the KDA resolution is not well-constrained, however, these distance estimates may be substantially different and $\overline{d}$ is not a good estimator of the distance. The uncertainty associated with $\overline{d}$ may be computed from the second moment of the DPDF as

$$\sigma_{\bar{d}} = \left(\int_0^\infty d_\odot^{\,2} \ \mathrm{DPDF} \ \mathrm{d}(d_\odot) \ - \bar{d}^{\,2} \right)^{1/2} \ . \tag{2.2}$$

The $\sigma_{\bar{d}}$ represent the variance of the DPDF, and only approximate Gaussian confidence intervals for single-peaked DPDFs. Ultimately, the choice of a single-value distance estimate will depend on the specifics of the application; various cases are discussed in §2.6.1.2.

2.3.3 Using DPDFs to Estimate Physical Parameters

While distances to objects are often interesting in isolation, their primary use is to convert observational quantities into physical properties of the object. DPDFs offer a simple way to propagate the uncertainties in distance through these calculations. For example, the maximum-likelihood mass of a molecular cloud clump can be estimated as

$$M_{\rm ML} = \alpha \; S_{1.1} \; d_{\rm ML}^{\,2} \ , \tag{2.3}$$

where $S_{1.1}$ is the $\lambda = 1.1$ mm flux density, and α contains the dust physics and temperature. Adoption of a DPDF representation allows marginalization over distance to obtain the expectation value of the mass:

$$\langle M \rangle = \int_0^\infty \alpha \; S_{1.1} \; d_\odot^{\,2} \; \mathrm{DPDF} \; \mathrm{d}(d_\odot) \ . \tag{2.4}$$

Practically, this integration can be accomplished by Monte Carlo methods, drawing a large number of distance samples from the DPDF and evaluating the average mass. Uncertainties in the expectation value can be determined using methods paralleling those used for distance above.

Bimodal DPDFs again lead to complications, as the expectation value will commonly be found at a value with low probability. A maximum likelihood distance can be adopted to avoid this aesthetic feature, but marginalization over the distance remains the most rigorous approach. Ideally, additional prior DPDFs should be applied in order to minimize bimodality.

2.3.4 Kinematic Distance DPDFs

Kinematic distances form the foundation for the Bayesian approach to distance estimation, computed from the intersection of the Galactic rotation curve projected along the line of sight, $v(d_\odot)$, with the observed molecular line $v_{\rm LSR}$. Transformation of velocity uncertainties onto the distance axis is facilitated by the use of two-dimensional probability density functions, $P(v_{\rm LSR}, d_\odot)$.

The rotation curve function, $P_{\rm rotc}(v_{\rm LSR}, d_\odot)$, is constructed as

$$P_{\rm rotc}(v_{\rm LSR}, d_\odot) = \exp\left(- \frac{\left[v_{\rm LSR} - v(d_\odot)\right]^2}{2\sigma_{\rm vir}^2} \right) \ , \tag{2.5}$$

where the uncertainty $\sigma_{\rm vir}$ 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^{-1}; Reid et al., 2009b)[5] . The function is Gaussian in $v_{\rm LSR}$, and is centered along $v(d_\odot)$; if integrated over $v_{\rm LSR}$, a uniform DPDF is obtained. The probability density function from spectral line information ($P_{\rm spec}$) is a Gaussian centered at the measured $v_{\rm line}$, with observed linewidth $\sigma^2_{\rm line}$, independent of $d_\odot$. As with $P_{\rm rotc}$, this function yields a uniform DPDF when integrated over $v_{\rm LSR}$. Since $P_{\rm rotc}$ does vary as a function of $d_\odot$, localized peaks in the ($v_{\rm LSR}$, $d_\odot$) plane result when it is multiplied by $P_{\rm spec}$. The desired one-dimensional DPDF$_{\rm kin}$ is obtained by subsequent integration over $v_{\rm LSR}$.

DPDF$_{\rm kin}$ is double-peaked and symmetric about the tangent distance for objects with $R_{\rm gal} < R_0$, and single-peaked otherwise. The $v(d_\odot)$ were computed using the flat rotation curve of Reid et al. (2009b). Schönrich et al. (2010) subsequently derived newer estimates of the Solar peculiar motion, affecting rotation curve fits to the maser parallax data of Reid et al. The updated values used here are R_0 = 8.51 kpc, and Θ_0 = 244 km s^{-1} (M. Reid 2011, private communication). The new solar motion values also had the effect of decreasing the magnitude of the apparent Galactic counter-rotation of high-mass star forming regions, an effect likely arising from molecular gas interacting with the spiral potential, from 15 km s^{-1} to 6 km s^{-1}.

Kinematic distances are sensitive to the slope of $v(d_\odot)$, itself a function of Galactic longitude. For lines of sight along $b \approx 0°$ within $\sim$ 10° of the Galactic longitude cardinal directions, $v(d_\odot)$ is either very flat or sharply peaked; small departures from circular motion therefore translate into large deviations in derived kinematic distances. Furthermore, since $v(d_\odot)$ is derived assuming circular orbits about the Galactic center, radial streaming motions of the gas are not accounted for, meaning that DPDF-derived distance estimates carry the basic limitations of any kinematic distance determination. To minimize the effects of non-circular motion, regions known to have significant streaming must be excluded from consideration. In particular, the presence of the long Galactic bar at $R_{\rm gal} \lesssim 3$ kpc (Fux, 1999; Rodriguez-Fernandez & Combes, 2008) and its associated radial streaming motions restrict the use of kinematic distance measurements to locations outside this radius. In the Galactic longitude-velocity ($\ell - v$) diagram, these restrictions amount to excluding

[5] This is the expected virial velocity, per coordinate, for an individual object (**i.e.**, molecular cloud clump) within a high-mass star-forming region of mass $\sim 3\times10^4$ $M_\odot$ and radius $\sim$1 pc (Reid et al., 2009b).

much of $|\ell| \lesssim 20°$. Features at low longitude known to be outside the Galactic bar (such as the Scutum-Centarus arm, also labeled as the "Molecular Ring"; Dame et al., 2001, their Fig. 3), may be considered to have roughly circular orbits, and are included in this study.

2.3.5 Prior DPDFs for Kinematic Distance Discrimination

Prior DPDFs are required to discriminate between the kinematic probability peaks for objects within the solar circle. $\mathrm{DPDF_{kin}}$ is symmetric about the tangent point, so prior DPDFs based on ancillary Galactic plane data must be asymmetric to provide useful distance constraints.

The Galactic distribution of molecular gas serves as an envelope inside which molecular cloud clumps may form. The prior $\mathrm{DPDF_{H_2}}$ is defined to be proportional to the volume density from the molecular hydrogen model of Wolfire et al. (2003) along a line of sight. This model consists of a Molecular Ring component with a decaying exponential toward the outer Galaxy; the vertical distribution is Gaussian with a half-width at half maximum of 60 pc (Bronfman et al., 1988), flaring outside the Solar Circle. While this distribution is symmetric about d_{tan} along the Galactic midplane, the narrow vertical extent of the molecular layer sets a strong prior on higher-latitude objects. The relative amount of H_2 beyond the tangent point for lines of sight at $|b| \gtrsim 0.^\circ 3$ is small, generating the needed asymmetric function for molecular cloud clumps at larger Galactic latitude.

The prior DPDF based on EMAFs was computed from a pixel-by-pixel morphological matching between millimeter dust-continuum emission and mid-infrared dust absorption features. The derivation of $\mathrm{DPDF_{emaf}}$ is described in detail in the next section.

2.4 Infrared-Millimeter Morphological Matching

Morphological matching is based on the comparison between synthetic 8-μm images computed from millimeter flux density measurements and GLIMPSE 8-μm maps processed to match the angular resolution of the BGPS. This section describes the creation of both the synthetic and processed 8-μm images, as well as the mechanics of computing $\mathrm{DPDF_{emaf}}$.

2.4.1 Creation of Synthetic 8-μm Images

2.4.1.1 Radiative Transfer Assumptions

Creation of synthetic 8-μm images explicitly assumes that the dust seen in emission in the BGPS is the same dust that extincts mid-infrared light. When converted into a mid-infrared optical depth, BGPS observations represent dark clouds which may be placed at different heliocentric distances within a model of diffuse Galactic 8-μm emission. A series of synthetic images generated in this manner were compared with mid-infrared observations to compute the $\mathrm{DPDF_{emaf}}$.

We assumed a simple radiative transfer model to describe the observed mid-infrared intensity absorbed by a cold molecular cloud clump immersed in a sea of diffuse emission (assuming that the absorbing cloud has no emission). The intensity observed within an EMAF (I_{emaf}) is

$$I_{\mathrm{emaf}} = I_{\mathrm{back}}\ e^{-\tau_8} + I_{\mathrm{fore}}\ , \tag{2.6}$$

where I_{back} and I_{fore} are the background (from the cloud to large heliocentric distance) and foreground (between the observer and the cloud) intensities, respectively, and τ_8 is the mid-infrared optical depth of the cloud. The total intensity along a line-of-sight in the absence of absorption is $I_{\mathrm{MIR}} = I_{\mathrm{back}} + I_{\mathrm{fore}}$. Defining the fraction of the total intensity that lies in front of the cloud as $f_{\mathrm{fore}} = I_{\mathrm{fore}}/I_{\mathrm{MIR}}$ allows Equation (2.6) to be written as

$$I_{\mathrm{emaf}} = \left[(1 - f_{\mathrm{fore}})\ e^{-\tau_8} + f_{\mathrm{fore}}\right]\ I_{\mathrm{MIR}}\ . \tag{2.7}$$

This parameterization frames the observed EMAF intensity in terms of decrements below the unextincted intensity in the vicinity, and provides the basis for creating synthetic 8-μm images. It follows quickly from Equation (2.7) that clouds optically thick in the mid-infrared ($\tau_8 \sim 1$) will still have a 10% difference between I_{emaf} and I_{MIR} (**i.e.**, easily detectable) for f_{fore} as large as 0.85. Calculation of τ_8 and f_{fore} are described below, and the estimation of I_{MIR} from GLIMPSE data is discussed in §2.4.2.

2.4.1.2 8-μm Optical Depth from the Millimeter Flux Density

The mid-infrared optical depth of an EMAF cannot be measured directly from the GLIMPSE mosaics without significant assumptions, but it may be estimated from millimeter data. Thermal dust emission is optically thin at millimeter wavelengths, so the observed BGPS flux density ($S_{1.1}$) may be written as

$$S_{1.1} = B_{1.1}(T_d)\ \tau_{1.1}\ \Omega_{\rm BGPS}\ , \tag{2.8}$$

where $B_{1.1}(T_d)$ is the Planck function evaluated at $\lambda = 1.1$ mm and dust temperature T_d, and $\Omega_{\rm BGPS} = 2.9 \times 10^{-8}$ sr is the solid angle of the BGPS beam. The millimeter optical depth ($\tau_{1.1}$) was computed assuming the dust opacity ($\kappa_{1.1}$) for grains with thin ice mantles, coagulating at 10^6 cm^{-3} for 10^5 years (Ossenkopf & Henning, 1994, Table 1, Column 5; called OH5 dust). Interpolation of OH5 dust opacities to the central frequency of the BGPS bandpass yields $\kappa_{1.1} = 1.14$ cm^2 g^{-1} of dust (A11). A molecular cloud clump with $\tau_{1.1} = 10^{-3}$, which corresponds to ($S_{1.1} \approx 0.9$ Jy), has a beam-averaged molecular hydrogen column density $\approx 2 \times 10^{22}$ cm^{-2}.

The 8-μm optical depth is related to $\tau_{1.1}$ by the ratio of the dust opacities in the two bandpasses, $R_\kappa = \kappa_8/\kappa_{1.1}$. We calculated the mid-infrared dust opacity by assuming a dust emission spectrum including PAH molecules (Draine & Li, 2007), finding the average attenuated intensity across IRAC Band 4, and extracting a band-averaged opacity $\kappa_8 = 825$ cm^2 g^{-1} of dust (see Appendix 2.8.1). At the 33$''$ resolution of the BGPS, the beam-averaged 8-μm optical depth is therefore

$$\begin{aligned}\tau_8 &= \frac{R_\kappa}{B_{1.1}(T_d)\ \Omega_{\rm BGPS}}\ S_{1.1} = \Upsilon(T_d)\ S_{1.1} \\ &= 0.778\ \left(\frac{e^{13.0{\rm K}/T_d} - 1}{e^{13.0{\rm K}/20.0{\rm K}} - 1}\right)\left(\frac{S_{1.1}}{1\ {\rm Jy}}\right)\ .\end{aligned} \tag{2.9}$$

The function $\Upsilon(T_d)$ has units of inverse flux density, and is normalized to 20 K in Equation (2.9). Because τ_8 is a function of R_κ (**i.e.**, both millimeter-wave emission and mid-infrared absorption depend only on the dust), the dust-to-gas ratio is not relevant to the distance estimation method. Owing to the nearly three orders of magnitude difference in dust opacity between the millimeter and mid-infrared, a value of $\tau_8 = 0.1$ corresponds to a column of only $N(\mathrm{H_2}) \approx 3 \times 10^{21}$ cm^{-2}, assuming

$A_{[8\mu]}/A_V \approx 0.05$ (Indebetouw et al., 2005; Román-Zúñiga et al., 2007). Therefore, molecular cloud clumps with column densities $\gtrsim 10^{22}$ cm^{-2} will be mostly opaque at $\lambda = 8$ μm.

Using Equation (2.9) to obtain an 8-μm optical depth requires a dust temperature (T_d). Since we are ignorant of T_d within each molecular cloud clump used in this study, we assumed that all sources are at the same temperature. Battersby et al. (2011) showed that mid-infrared-dark molecular cloud clumps generally span the temperature range 15 K $\lesssim T_d \lesssim$ 25 K. Therefore, $T_d = 20$ K is a reasonable representation for BGPS sources as a group. Variation of the assumed T_d affects the KDA resolutions for some sources, and is discussed briefly in §2.6.1.1. With molecular cloud clump dust temperatures derived from **Herschel** Hi-GAL data, more precise DPDFs for individual objects may be derived using the present methodology.

2.4.1.3 8-μm Foreground Fraction from a Galactic Emission Model

Absorption features seen at $\lambda = 8$ μm are assumed to be the result of dense clouds immersed in a smooth emission distribution, punctuated by regions undergoing active star formation. While small-scale structures are difficult to model, the broader diffuse emission is a more tractable problem. Creation of synthetic 8-μm images via Equation (2.7) requires a three-dimensional model for the Galactic 8-μm emission distribution.

The recent numerical Galactic stellar and dust emission model of Robitaille et al. (2012, hereafter R12), computed using the Monte-Carlo three-dimensional radiative transfer code Hyperion[6] (Robitaille, 2011), offers a self-consistent estimate of diffuse Galactic emission that is well-matched to observed quantities. We used the final model presented in R12, whose parameters were chosen to fit the Galactic latitude and longitude intensity distributions from seven bandpasses in the mid- to far-infrared. This model features two major and two minor spiral arms with Gaussian radial profiles, a lack of dust in the inner few kiloparsecs of the Galactic disk (dust hole; correlated with the dearth of molecular gas in this region), and a modified PAH abundance relative to the favored model from Draine & Li (2007). An analysis of the contributions from various stellar populations

[6] `http://www.hyperion-rt.org`

Table 2.2. Comparison of Hyperion Model Parameters

Category	Parameter	R12	This Work
Grid[a]	N_R	200	200
	N_ϕ	100	200
	N_z	50	44
	$\lvert z\rvert_{\rm max}$ (pc)	3000	1000
Wavelength[b]	N bins	160	22
	Range (μm)	$3 \leq \lambda \leq 140$	$6 \leq \lambda \leq 10$
Image[c]	Observer $R_{\rm gal}$ (kpc)	8.5	8.5
	Observer z (pc)	+15	+25
	Longitude Range (°)	$65 \geq \ell \geq -65$	$65 \geq \ell \geq -65$

[a] N = number of grid cells in this dimension

[b] Wavelengths at which the model images were computed later convolved with instrument bandpasses to create simulated observations.

[c] Parameters related to observer within the grid.

and dust grain sizes to the total intensity in each bandpass indicates that some 96% of the emission detected in IRAC Band 4 images comes from PAH molecules (R12).

A three-dimensional ($\ell, b, d_\odot$) data cube of 8-μm emission was generated from the radiative transfer code using model grid and image parameters slightly modified from those used by Robitaille et al. Table 2.2 lists the comparison of Hyperion input parameters between R12 and the present study. Primary differences include an increase in azimuthal resolution of the cylindrical grid, a restriction of the vertical extent of the grid to $|z| \leq 1$ kpc (to match the region of the Galactic plane probed by the latitude range $|b| \leq 1°$), and limiting the wavelength range used in computing output images. The model was computed within a box 30 kpc on a side, containing the entire modeled stellar disk (R12); a face-on view of the model Milky Way as seen from the north Galactic pole is shown in Figure 2.1. Lines of sight out to 20 kpc (the distance used for DPDF generation) lie entirely within the simulation box for $|\ell| \leq 48°$. Beyond this longitude, however, the edge of the box retreats to only $d_\odot \approx 16.5$ kpc by $|\ell| = 65°$.

A series of (ℓ, b) images of the Galactic plane containing only emission between the observer and some distance d_i were computed using a useful Hyperion feature, using a resolution of $3'$

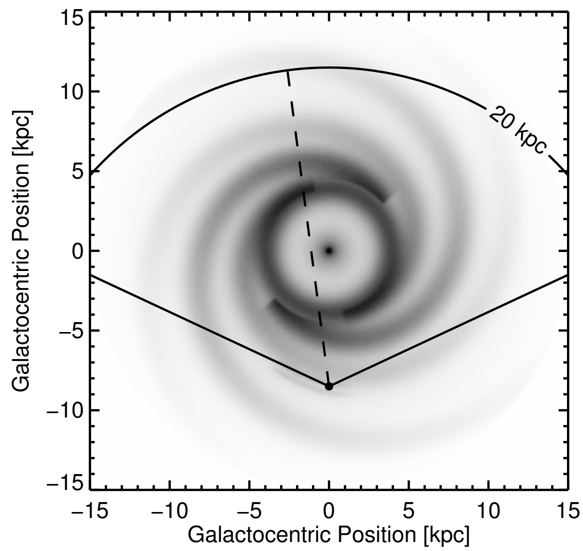

Figure 2.1 Galactic mid-infrared emission model computed with Hyperion (Robitaille et al., 2012) viewed from the North Galactic Pole. The model, viewed through the IRAC 8.0-μm bandpass, is shown on an inverted square-root intensity scale. The Sun is located at $(x, y) = (0, -8.5$ kpc), and solid diagonal lines represent the limits of the GLIMPSE survey ($|\ell| = 65°$). The dashed line marks the low-latitude ($\ell = 7\overset{\circ}{.}5$) limit of this study. DPDFs were computed out to $d_\odot = 20$ kpc (curved contour).

in latitude, and 15′ in longitude for computational reasons. Images were computed for each of 22 wavelength bins logarithmically spaced from 6 to 10 μm (closely matching the wavelength bins used by R12 for this part of the spectrum), and were then convolved with the IRAC Band 4 transmission curve to yield a single simulated **Spitzer** image (see Robitaille et al., 2007 for complete details). The three-dimensional image was constructed by stepping d_i outward in 100-pc intervals and depicts the cumulative 8-μm emission out to each d_i. The resulting cube comprises 200 steps, with the final image slice including all model emission to the edge of the box, equivalent to the collapsed profiles presented in R12.

Hyperion cannot treat sources individually, but rather uses "diffuse" sources of emission in each grid cell. These diffuse sources are generated from the probability of emission from various populations of stars and similar objects (**e.g.**, planetary nebulae, H II regions), assigned a spectrum corresponding to the appropriate spectral class, and given a total luminosity based on the number of "real" sources the cell represents. While most of the emitting populations have a smooth spatial

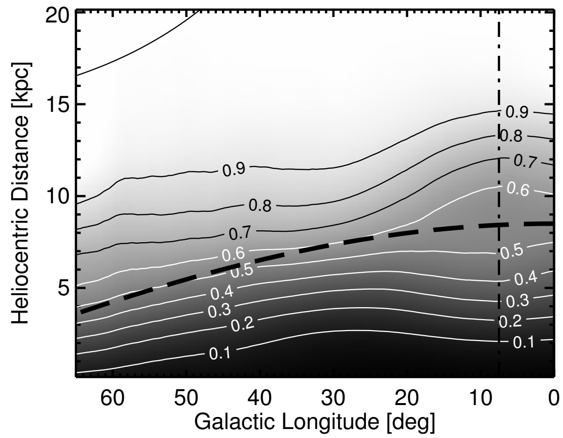

Figure 2.2 Foreground fraction of Galactic 8-μm emission in the northern Galactic plane derived from the HYPERION model as a function of (ℓ,$d_\odot$) along $b = 0°$. Grayscale and contours represent $f_{\rm fore}$, with the unlabeled 1.0 contour marking the edge of the box in Fig. 2.1 for $\ell \gtrsim 48°$. The thick black dashed line follows the tangent distance as a function of Galactic longitude, and the vertical dot-dashed line marks the $\ell = 7\overset{\circ}{.}5$ lower limit of this study.

distribution, relatively rare sources with concentrated emission at $\lambda = 8$ μm (such as H II regions) are sprinkled throughout the box according to the underlying stellar distribution model. Very nearby objects ($d_\odot \leq 0.5$ kpc) appear quite bright, and cause "hot-pixel" effects in the computed images of the Galactic plane. These objects blend into the background for images computed from large Galactocentric position (**e.g.**, Fig. 2.1), or are averaged out in collapsed longitude or latitude distributions (R12). To ameliorate the effect of these objects in the computed (ℓ, b) images, we ran seven realizations of the model, each with a different random-number seed, then median-combined the realizations of each d_i slice. Since the underlying distribution of sources is fixed, nearby bright sources often appear in the same pixel in the output images; the number of realizations was chosen to be large enough such that median combining the realizations removes most of these outliers. To eliminate any remaining outliers and reduce noise, the combined (ℓ, b) images were median smoothed with a 3 pixel $\times$ 3 pixel box.

The foreground fraction was computed from the intensity cubes by dividing each (ℓ, b) image slice by the final slice. The final FITS data cubes of 8-μm intensity and $f_{\rm fore}$ for both the northern and southern Galactic plane ($|\ell| \leq 65°$) are publicly available with the BGPS archive. To illustrate the Galactic features present in the modeled cube, $f_{\rm fore}(\ell,d_\odot)$ for the Northern plane along $b = 0°$ is shown in Figure 2.2, with contours and grayscale representing its value from 0 to 1. Since PAH molecules contribute the bulk of the model emission, the dust hole towards low longitude is visible as a flattening of $f_{\rm fore}(d_\odot)$. The Molecular Ring / Scutum tangent at $\ell \approx 30°$ appears where $f_{\rm fore}$ grows quickly as a function of distance. The limited distance range caused by the model box size is represented by the 1.0 contour for $\ell \gtrsim 48°$. The tangent distance as a function of longitude (black dashed line) spans the range $0.45 \lesssim f_{\rm fore} \lesssim 0.6$, implying that clouds that are optically thick in the mid-infrared should be visible beyond $d_{\rm tan}$.

2.4.1.4 Computing the Synthetic Images

Synthetic images ($I_{\rm emaf}$) for a given BGPS object are computed using Equation (2.7). The optical depth is modeled as a two-dimensional image, constructed by applying Equation (2.9) to the BGPS map data. The estimate of the total mid-infrared emission ($I_{\rm MIR}$) is also a two-dimensional image, and its creation is discussed below. Because of the coarse resolution of the $f_{\rm fore}$ model, we simply extracted the one-dimensional $f_{\rm fore}(d_\odot)$ at the (ℓ, b) of the BGPS object. The combination of these elements yields a cube of synthetic data to be compared with the processed GLIMPSE images.

2.4.2 Processing of GLIMPSE 8-μm Images

Mid-infrared properties of dust-continuum-identified molecular cloud clumps were derived from the **Spitzer**/GLIMPSE mosaics. Further processing of these images was required to estimate the total mid-infrared intensity ($I_{\rm MIR}$) in the vicinity of an EMAF, and to produce a smoothed, star-subtracted map, containing features and angular scales comparable to (sub-)millimeter data. The example source G035.524-00.274 (BGPS #5647) is used to illustrate the processing products in Figure 2.3. The first step was to remove individual stars because they contaminate estimates of broader diffuse emission and (sub-)millimeter observations are not sensitive to them. Star locations were identified by searching for bright, unresolved objects in the 3.6-μm mosaics using `DAOFIND`

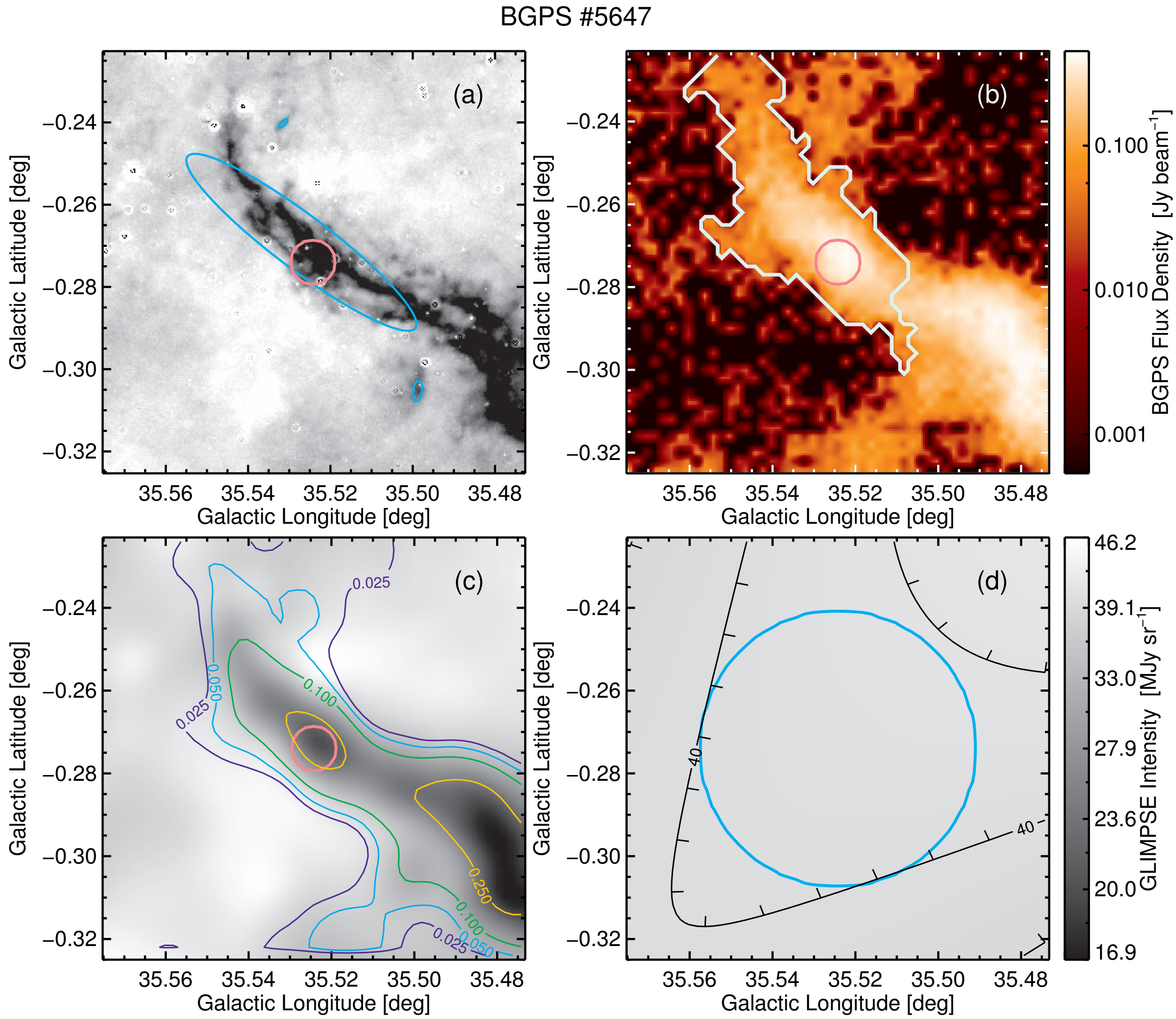

Figure 2.3 BGPS and processed GLIMPSE data for example object G035.524-00.274. All panels are $6' \times 6'$ postage-stamp images (see text), and the pink circle identifies the $40''$ top-hat BGPS equivalent aperture, centered on the location of peak flux density. (**a**) Cutout of the star-subtracted GLIMPSE image at native resolution. Cyan ellipses mark IRDCs identified in the Peretto & Fuller (2009) catalog; note that the dark cloud associated with BGPS source G035.478-00.298 (lower-right corner) is not included in that catalog. (**b**) BGPS map data with Bolocat source boundary (light cyan). (**c**) Star-subtracted GLIMPSE cutout smoothed to $33''$ and resampled to $7''\!.2$ pixels to match the BGPS maps. Color contours represent logarithmic flux density levels from BGPS (Jy beam^{-1}). (**d**) Estimate of the total mid-infrared intensity, $I_{\rm MIR}$, as a quadratic surface fitted to background pixels as described in the text. Contours are drawn to show the variation in $I_{\rm MIR}$ over the postage stamp (MJy sr^{-1}; ticks point to higher values), and the cyan aperture marks the region used to estimate $\langle I_{\rm MIR}\rangle$ for this source (see §2.5.1). The grayscale colorbar represents the common (logarithmic) intensity scale for all three GLIMPSE panels. (A color version of this figure is available in the online journal.)

(Stetson, 1987; Landsman, 1995) with a threshold of 20 MJy sr^{-1}. A Gaussian was fit to the 8-μm image at the location of each identified star, then subtracted. This method of star subtraction was deemed optimal because PSF variations across the survey mosaics meant that PSF-based approaches could not be applied. Star subtraction in this manner did, however, leave clear low-level residuals (Fig. 2.3**a**). Since later processing smooths the resulting images to the BGPS resolution, residuals are largely unimportant. However, to ensure that poor star subtraction or other effects did not effect distance estimation, by-eye evaluation of each potential EMAF for contamination was performed.

For further processing of the GLIMPSE data, $6' \times 6'$ postage-stamp images were extracted from the star-subtracted 8-μm mosaics for each Bolocat source. These postage stamps, centered on the location of peak millimeter flux density, limit consideration of mid-infrared variations to the immediate vicinity of a molecular cloud clump in addition to providing computational expediency. The first postage-stamp image created for a given BGPS object is a version of the star-subtracted GLIMPSE mosaic, re-pixelated and aligned to the $7\farcs2$ scale of the BGPS images. This image was used to ensure that locally bright emission did not interfere with derived mid-infrared intensities, and to determine the likely intensity range containing $I_{\rm MIR}$ around the object. A pixel intensity histogram of the image was constructed with 1 MJy sr^{-1}-wide bins, and the background was defined as intensities within its full-width at half maximum. Pixels within an $8' \times 8'$ section of the native-resolution star-subtracted GLIMPSE mosaic having intensities in the defined range were used to fit a quadratic surface using a linear, least-squares optimization. This surface, repixelated and scaled as above, comprises the postage-stamp estimate of $I_{\rm MIR}$ (Fig. 2.3**d**). This estimate of background pixels ignored high pixel values from star residuals and low pixel values from EMAFs.

The IRAC camera on **Spitzer** suffers from internal scattering of light which affects instrument calibration (Reach et al., 2005). Point-source photometry is unaffected by the scattering due to the calibration technique employed, but extended emission (such as the Galactic plane) will appear brighter due to scattering into each pixel. Correcting for this effect should be done on a frame-by-frame basis, but was not accounted for in the GLIMPSE pipeline (S. Carey 2010, private communication). To approximately correct for the scattering, an estimate of the scattered light was subtracted from the postage-stamp images for each BGPS source. The postage-stamp size was chosen to be near the $5\farcm2 \times 5\farcm2$ FOV of IRAC, and the $I_{\rm MIR}$ fit serves as the estimate of the light available to be scattered within a single IRAC frame. This estimate is only approximate, as the $I_{\rm MIR}$ fit explicitly excludes very bright and very dim emission within a frame; for frames with regions of bright emission, the derived correction factor will be a lower limit, and vice verse for frames containing extensive dark clouds. The infinite-aperture intensity correction for IRAC Band 4 is 0.737 (Reach et al., 2005), meaning that $\xi = 0.263$ is the scattered light fraction. Assuming that $I_{\rm MIR}$ represents the total incident light, we subtracted $\xi \times \mathrm{median}(I_{\rm MIR})$ from each postage-stamp image to remove scattered light.

Reduction of the GLIMPSE angular resolution was necessary for direct comparison with the synthetic images created using Equation (2.7). Since bright emission in the vicinity is scattered into an EMAF, removal of the scattered light must be done prior to smoothing. The scattering-corrected extracted postage stamps were smoothed with a FWHM = $33''$ Gaussian kernel, then re-pixelated and aligned to match the BGPS images (Fig. 2.3**c**).

2.4.3 Morphological Matching

Derivation of $\mathrm{DPDF}_{\rm emaf}$ relies upon the comparison of the (sub-)millimeter emission and mid-infrared absorption of dust in cold molecular cloud clumps. The series of synthetic images were matched against the smoothed GLIMPSE postage stamp images (Fig. 2.3**c**). For small $d_\odot$, the synthetic cloud appears darkest, without foreground light filling in the absorption feature. At larger heliocentric distance, $f_{\rm fore}$ grows, and the synthetic image converges upon $I_{\rm MIR}$ (Fig. 2.3**d**).

The Galactic 8-μm emission model of §2.4.1.3 describes smooth, diffuse emission against which EMAFs are visible, but actual Galactic emission is more complex. To match more closely the assumption of the model, the angular region over which the synthetic and observed sky are compared must be restricted. As the observational definition of a single molecular cloud clump, we began with a source's Bolocat contour delineating the maximum extent of the comparison region

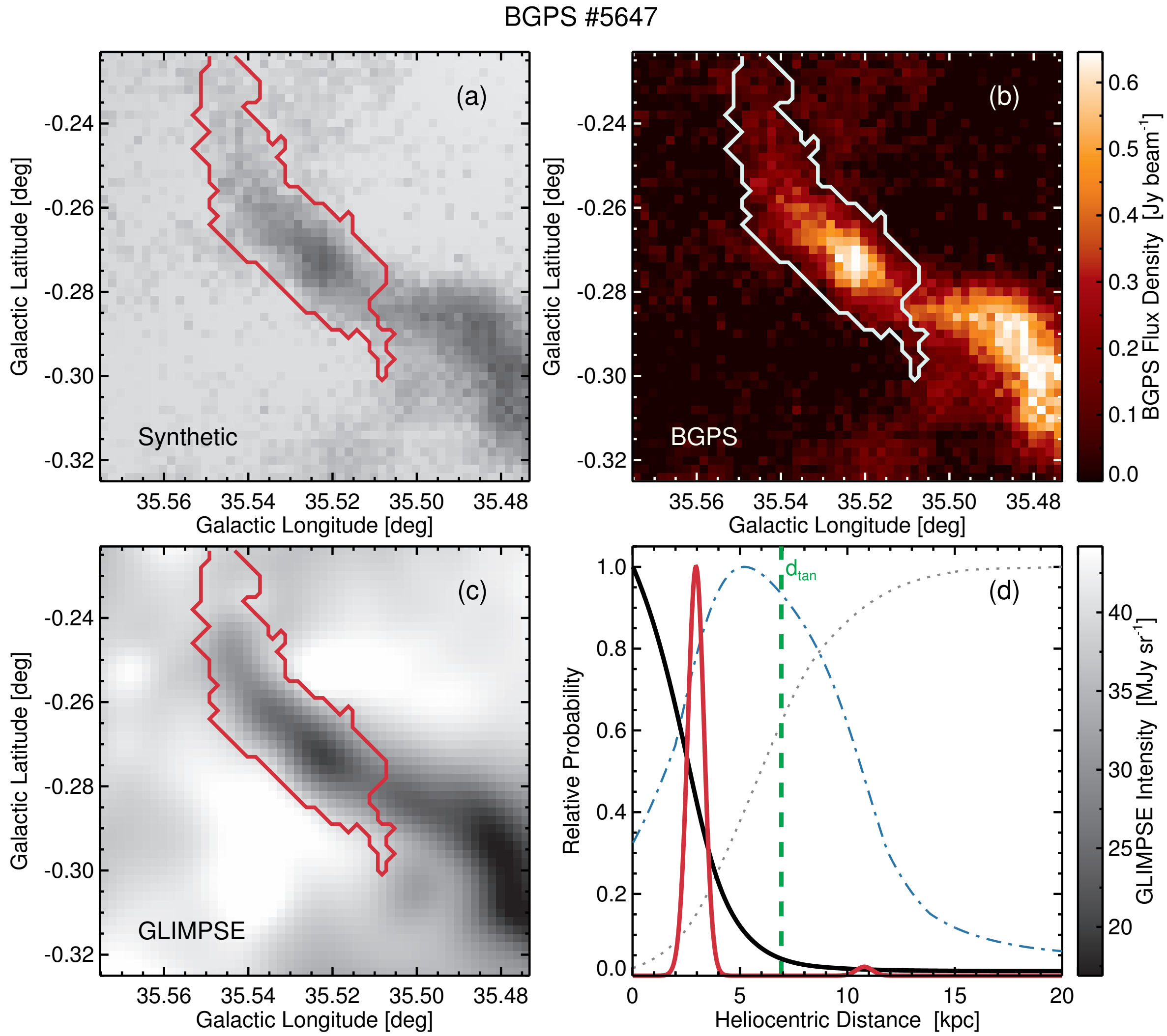

Figure 2.4 Morphological matching for example object G035.524-00.274. The millimeter source to the lower-right of the marked contour is a separate Bolocat object. (**a**) Synthetic 8-μm image calculated via Equation (2.7). (**b**) BGPS postage-stamp image, showing the restricted region used for the morphological matching (see text). (**c**) Smoothed GLIMPSE map against which the synthetic images are compared. (**d**) Prior DPDFs from the morphological matching (black) and molecular gas distribution (blue dot-dashed), and the posterior DPDF (red), including the kinematic distance likelihood (see text). The gray dotted line represents f_{fore} extracted from the model cube along (ℓ, b), and the green dashed line marks the tangent distance. (A color version of this figure is available in the online journal.)

(Rosolowsky et al., 2010b). Since the synthetic image can never be brighter than I_{MIR}, the matching process is adversely affected by bright mid-infrared emission in the vicinity of a BGPS source. To ameliorate this effect, pixels in the smoothed GLIMPSE postage stamp were excluded from the matching region if their value exceeded the corresponding value in the $I_{\rm MIR}$ image.

An overview of the morphological matching process is presented in Figure 2.4 for the same object as in Figure 2.3. The synthetic 8-μm image is shown in panel (**a**) for the distance which maximizes DPDF$_{\rm emaf}$ (see below). Panels (**b**) and (**c**) are identical to Figure 2.3, except that the source contour now marks the restricted matching region due to bright mid-infrared emission on the perimeter of the EMAF. Panels (**a**) and (**c**) are shown on a common linear grayscale to illustrate the match between the observed extinction and that predicted from thermal dust emission. The various DPDFs for this source are shown in panel (**d**), and are described below.

Quantification of the match as a function of distance was accomplished by constructing a χ^2 statistic from a pixel-by-pixel comparison within the matching region. The estimate of the error in each pixel was derived from Equation (2.7) by propagating the uncertainty in the optical depth map as

$$\sigma_{\rm syn}(\ell,b) = I_{\rm MIR}(\ell,b)\ e^{-\tau_8(\ell,b)}\ \Upsilon\ \sigma_{{\rm S}_{1.1}}\ , \tag{2.10}$$

where Equation (2.9) defines τ_8 and Υ, and $\sigma_{{\rm S}_{1.1}}$ is the median absolute deviation of the BGPS postage-stamp image. This estimate of the uncertainty places more weight on the portions of the image with larger BGPS flux density. The statistic was computed for synthetic images at 100-pc intervals along the line of sight, yielding $\chi^2(d_\odot)$.

A preliminary DPDF$_{\rm emaf}$ was computed using the formal probability of the $\Delta\chi^2$ statistic. The number of degrees of freedom was taken as the integer number of BGPS beams in the matching region ($N_{\rm pixels}$ / 23.8 pixels beam^{-1}; A11) minus one, since only beam-scale structures are independent and distance is a fitted parameter. For most sources, the DPDF$_{\rm emaf}$ has a broad peak (several kiloparsecs wide), and falls sharply where the $\Delta(\chi^2_{\rm red})$ exceeds unity. Because of the sharp cutoffs, it tends to very strongly favor one kinematic distance peak over the other. If the Galaxy truly consisted of dark molecular cloud clumps embedded within broad diffuse mid-infrared emission, this formulation of DPDF$_{\rm emaf}$ would be appropriate. However, the Galaxy is punctuated with regions of stronger 8-μm emission that violate the simple radiative transfer of Equation (2.6), and the $\mathrm{DPDF}_{\mathrm{emaf}}$ should contain a systematic uncertainty that allows non-vanishing probability at the non-favored kinematic distance peak.

Experimentation with alternative approaches that allow a systematic uncertainty led to the selection of $\mathrm{DPDF}_{\mathrm{emaf}} \propto (\chi^2)^{-\beta}$, where β is a positive scalar of order unity. This class of $\mathrm{DPDF}_{\mathrm{emaf}}$ have FWHM comparable to the formal probability, but greater width at low likelihood, and hence rarely goes to zero probability until far from the peak. The parameter β may be used to tune the width of the function, with larger values leading to narrower distributions. Since the sharp cutoff of the $\mathrm{DPDF}_{\mathrm{emaf}}$, not the width of the peak, is what appears problematic in light of complex Galactic emission, we selected $\beta = 2$ to reproduce the widths of the formal probability DPDF. To verify the validity of this choice, we computed the GRS distance matching success rate (see §2.5.3.2) as a function of β, and found no dependence on the width of $\mathrm{DPDF}_{\mathrm{emaf}}$.

The resulting DPDFs for object G035.524-00.274 (BGPS #5647) are shown in Figure 2.4**d**. The prior $\mathrm{DPDF}_{\mathrm{H_2}}$ (blue dot-dashed) favors the near kinematic distance, since this line of sight looks out the bottom of the molecular layer. The gray dotted line shows the $f_{\mathrm{fore}}(d_\odot)$ from the numerical model. The morphological matching process, represented by $\mathrm{DPDF}_{\mathrm{emaf}}$ (black solid), could not make the synthetic image dark enough to match the smoothed GLIMPSE image, forcing the prior to peak at $d_\odot = 0$ kpc. The posterior DPDF (red) clearly reflects the distance selection, with the near peak containing $> 95\%$ of the integrated probability, although there remains some probability contained in the far kinematic distance peak at $d_\odot \approx 11$ kpc.

2.5 Results

2.5.1 EMAF-Selected Molecular Cloud Clumps

2.5.1.1 Spatial and Kinematic Selection Criteria

We derived posterior DPDFs for the subset of BGPS sources that have a measured v_{LSR} from molecular spectroscopy, and are selected by the presence of an EMAF. Spatially, this set is defined by the GLIMPSE-BGPS overlap, limiting the upper end of the Galactic plane at $\ell = 65\overset{\circ}{.}25$ and

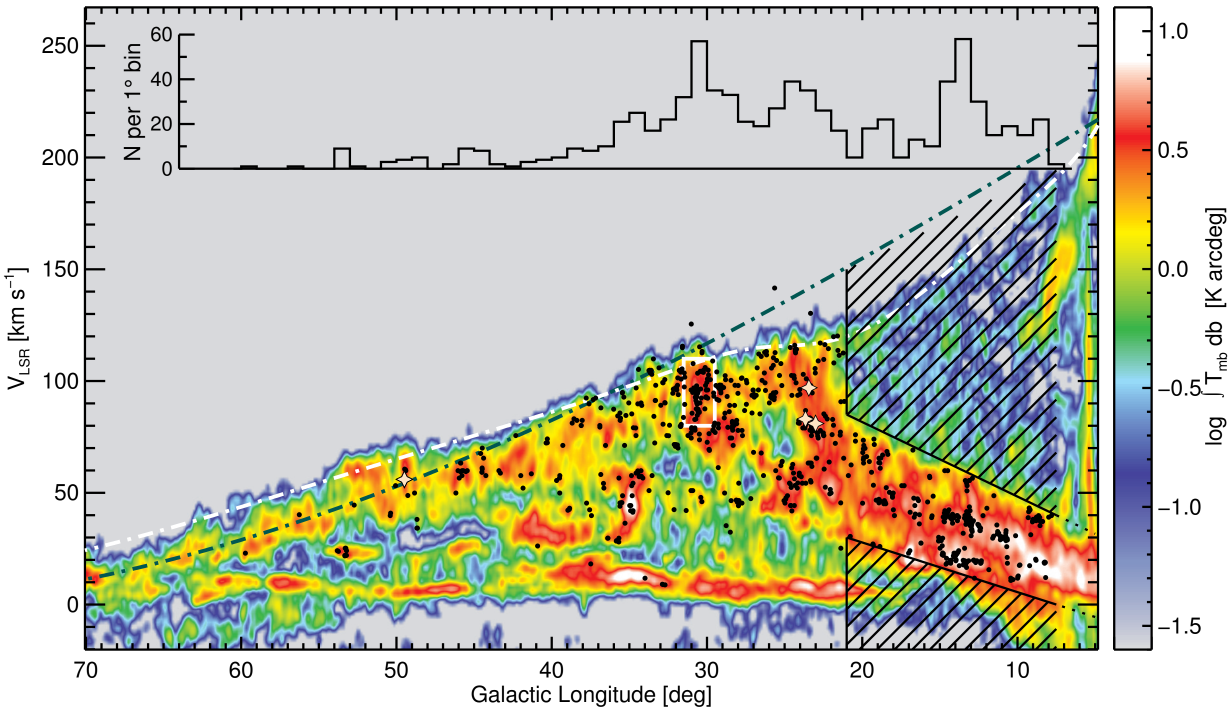

Figure 2.5 Longitude-velocity diagram of the northern Galactic plane. The background image is the latitude-integrated ($|b| \leq 2°$) ^{12}CO(1-0) intensity from Dame et al. (2001). Excluded regions, where the long Galactic bar causes significant deviations from circular motion, are hashed out. Black circles mark the locations of BGPS sources used in this study; the histogram at the top summarizes their Galactic longitude distribution. Stars show the locations of masers used for parallax distance comparison (see Table 2.4). The white rectangle at $(\ell, v_{\rm LSR}) \approx (30°, 100\ {\rm km\ s^{-1}})$ roughly marks the W43 star-formation region (see §2.6.1.1). Colored dot-dashed lines represent the Clemens (1985, white) and Reid et al. (2009b, dark green, includes counter-rotation of HMSFRs) rotation curve tangent velocities as a function of Galactic longitude. (A color version of this figure is available in the online journal.)

a latitude spread of $|b| \leq 1\overset{\circ}{.}0$. Kinematic considerations restrict the regions of the $\ell - v$ diagram (Fig. 2.5) that may be considered at lower Galactic longitude down to the $\ell = 7\overset{\circ}{.}5$ limit of the spectroscopic surveys.

The colored image in Figure 2.5 is the latitude-integrated ^{12}CO(1-0) intensity from Dame et al. (2001), and is shown as an indicator of molecular gas location and kinematic conditions. The presence of a long Galactic bar ($R_{\rm bar} \sim 4$ kpc; Benjamin et al., 2005) implies significant non-circular motion at $\ell \lesssim 30°$. Regions at these longitudes in the $\ell - v$ diagram associated with the Molecular Ring feature (**cf.** Dame et al., 2001; Rodriguez-Fernandez & Combes, 2008), however, are likely at $R_{\rm gal} \gtrsim 4$ kpc. To include the Ring but exclude bar-related gas, we disallowed the two hashed regions in Figure 5. The upper region is bounded by $v_{\rm LSR} = (3.33 \text{ km s}^{-1}) \times \ell(^\circ) + 15 \text{ km s}^{-1}$, and includes the higher-velocity gas inside the Ring. The lower region excludes the 3-kpc expanding arm, and is bounded by $v_{\rm LSR} = (2.22 \text{ km s}^{-1}) \times \ell(^\circ) - 16.7 \text{ km s}^{-1}$. Both regions are defined only for $\ell \leq 21^\circ$. The upper hashed region does not extend past this point because the Molecular Ring feature extends to the tangent velocity at larger longitudes; the lower region is limited because the 3-kpc arm has its tangency here (Dame & Thaddeus, 2008). There is likely overlap between Ring objects with nearly circular motions and objects in bar-related streaming orbits at $21^\circ \leq \ell \leq 30^\circ$, so kinematic distance estimates in this range, including those derived here, should be used with caution.

Black circles mark the locations of the BGPS molecular cloud clumps for which a DPDF$_{\rm emaf}$ was computed, and the histogram summarizes their longitude distribution. Stars mark the masers used for distance comparison (see §2.5.3.1).

2.5.1.2 Mid-Infrared Selection Criteria

Automated classification of dust-continuum-identified molecular cloud clumps as EMAFs was achieved using the mid-infrared contrast computed from the smoothed GLIMPSE images at the location of the BGPS source. The peak contrast was defined as

$$C = 1 - \frac{I_{\rm min}}{\langle I_{\rm MIR} \rangle} , \tag{2.11}$$

where the intensity values were measured from the processed postage-stamp images described in §2.4.2 for each Bolocat object. Due to the varied sizes and shapes of EMAFs, standardized intensities were measured in a $40''$ aperture around the location of peak BGPS flux density (pink circles in Fig. 2.3). The value of $I_{\rm min}$ is the minimum intensity within the aperture measured from the smoothed star-subtracted GLIMPSE image (Fig. 2.3**c**), and $\langle I_{\rm MIR} \rangle$ is the mean of the $I_{\rm MIR}$ postage-stamp image within $2'$ of the peak of millimeter flux density (cyan circle in Fig. 2.3**d**). A preliminary contrast threshold of $C \geq 0.01$ was implemented in the automated source selection to minimize the number of spurious matches caused by unrelated variation in the GLIMPSE 8-μm mosaics. This threshold also rejects BGPS sources that are mid-infrared bright, as those objects have negative contrast.

All molecular cloud clumps meeting the above selection criteria were examined by eye to ensure their suitability for deriving a DPDF$_{\rm emaf}$. Bolocat objects were not assigned a DPDF$_{\rm emaf}$ for the following types of deficiencies: (1) there was evidence of poor star subtraction contaminating $I_{\rm min}$; (2) there was very bright mid-infrared emission ($I_{8\mu{\rm m}} \gtrsim 200$ MJy sr^{-1}) within $2'$ of the location of peak BGPS flux density that could bleed into the $40''$ aperture or significantly affect the IRAC scattering correction; (3) the postage-stamp estimate of $I_{\rm MIR}$ was contaminated by excessive bright emission or dark extinction; or (4) the morphology of dark regions in the GLIMPSE image clearly did not correspond to that of the millimeter emission. By-eye exclusion removed approximately 40% of sources meeting the initial automated selection criteria.

Properties of rejected sources were analyzed to reveal that nearly all very-low contrast sources were spurious matches (deficiency type 4, see above). Additionally, BGPS objects located in fields of locally very bright mid-infrared emission were almost all excluded from the final source list (types 2 and 3). As a result, the contrast cutoff was increased to $C \geq 0.05$, and two additional automated selection criteria were introduced. First, the restriction $\langle I_{\rm MIR} \rangle \leq 100$ MJy sr^{-1} was placed to remove sources whose background estimate indicates significant disagreement with the 8-μm-emission model, as large discrepancies may lead to improper distance estimates (type 3). Second, to automatically reject sources near very bright emission, the re-pixelated unsmoothed postage-stamp images were checked for pixels with $I_{8\mu{\rm m}} \geq 200$ MJy sr^{-1} within $2'$ of the image center; sources with more than 10 such ($7\overset{\prime\prime}{.}2$) pixels were removed from consideration (type 2). These additions to the automated selection criteria led to fewer sources (only 28%) requiring by-eye removal, primarily due to poor star-subtraction (type 1) or complex emission structures that caused morphological mismatch (type 4).

2.5.1.3 Source Properties

Table 2.3. Observed & Derived Properties of EMAF-selected BGPS Molecular Cloud Clumps

BGPS V1.0 Catalog Properties[a]				Velocity Data						
Catalog Number	ℓ (°)	b (°)	S_{40}[b] (Jy)	$v_{\rm LSR}$ (km s^{-1})	Ref.	Mid-Infrared Contrast	KDA[c] Resol.	$P_{\rm ML}$[d]	$d_{\rm ML}$[e] (kpc)	$\overline{d}$[f] (kpc)
4638	30.990	0.329	0.186(0.048)	79.1	2	0.21(0.03)	N	0.88	$4.60^{+0.44}_{-0.40}$	···
4639	30.990	0.385	0.108(0.054)	78.7	2	0.08(0.03)	F	0.91	$9.88^{+0.38}_{-0.42}$	···
4650	31.016	-0.001	0.280(0.057)	74.5	1	0.25(0.04)	N	0.88	$4.46^{+0.42}_{-0.42}$	···
4653	31.026	-0.113	0.289(0.067)	76.8	1	0.40(0.03)	N	0.83	$4.50^{+0.52}_{-0.48}$	···
4655	31.032	0.783	0.267(0.110)	51.0	1	0.38(0.05)	N	0.99	$3.24^{+0.36}_{-0.36}$	···
4715	31.226	0.023	0.381(0.077)	74.5	1	0.40(0.02)	N	0.86	$4.40^{+0.48}_{-0.48}$	···
4749	31.342	-0.149	0.111(0.048)	42.0	1,3	0.15(0.03)	N	0.91	$2.84^{+0.44}_{-0.44}$	···
4769	31.432	0.167	0.117(0.039)	101.7	3	0.06(0.05)	U	0.51	···	···
4770	31.436	-0.103	0.122(0.039)	89.4	1,3	0.18(0.02)	U	0.70	···	···
4780	31.462	0.351	0.090(0.046)	97.3	3	0.09(0.01)	N	0.82	$5.56^{+0.72}_{-0.56}$	···
4781	31.466	0.185	0.255(0.054)	103.6	3	0.19(0.08)	U	0.70	···	···
4794	31.516	0.449	0.184(0.048)	83.7	3	0.11(0.04)	N	0.91	$4.94^{+0.42}_{-0.40}$	···
4811	31.580	0.227	0.186(0.048)	115.7	1,3	0.14(0.06)	T	0.57	$7.00^{+0.78}_{-0.62}$	7.13(0.66)
4814	31.584	0.205	0.218(0.052)	114.9	3	0.15(0.04)	T	0.61	$6.88^{+0.82}_{-0.56}$	7.07(0.66)
4826	31.608	0.171	0.120(0.042)	105.4	3	0.10(0.03)	U	0.60	···	···

[a]Rosolowsky et al. (2010b)

[b]Flux density and uncertainty within a 40″ aperture, corrected by the factor of 1.5 ± 0.15 from Aguirre et al. (2011)

[c]N = near; F = far; T = tangent point; U = unconstrained distance

[d]Integrated posterior DPDF on the side of $d_{\rm tan}$ containing $d_{\rm ML}$. Larger values indicate higher certainty in the KDA resolution.

[e]Maximum-likelihood distance; the distance where the posterior DPDF is largest. Not listed for unconstrained sources.

[f]Weighted-average distance; the first moment of the posterior DPDF. Only listed for sources at the tangent point.

References. — 1: HCO^+ (Shirley et al., 2013); 2: CS (Y. Shirley 2012, private communication); 3: NH_3 (Dunham et al., 2011b)

Note. — Errors are given in parentheses.

Note. — Table 2.3 is published in its entirety in a machine-readable format in the online journal. A portion is shown here for guidance regarding its form and content.

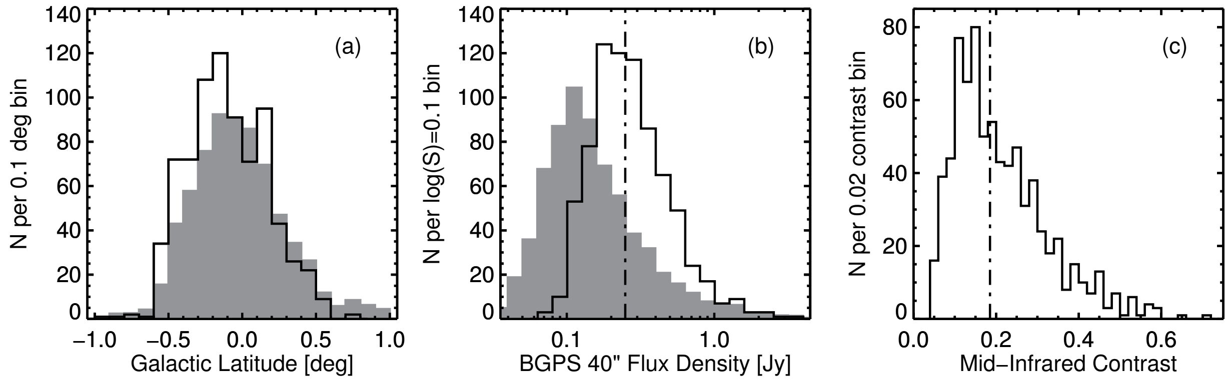

Figure 2.6 (**a**) Distribution of Galactic latitude of the sources in this study (black outline) and the entire BGPS catalog in the longitude range $7\overset{\circ}{.}5 \leq \ell \leq 65°$, divided by 10 (filled gray). The BGPS is nominally limited to $|b| \leq 0.5$. (**b**) Distribution of BGPS 40″ flux density for the sources in this study (black outline) and the full, longitude-limited BGPS catalog, divided by 10 (filled gray). Vertical dot-dashed line marks the median for this sample. (**c**) Distribution of measured mid-infrared contrast for the sources in Table 2.3. Vertical dot-dashed line marks the median.

The final source list contains 770 BGPS objects, and is presented in Table 2.3. EMAF-selected BGPS molecular cloud clumps are not drawn uniformly from the BGPS catalog. Comparisons of Galactic latitude and 40″ flux density distributions between this sample and the full Bolocat (within the spatial limits defined above) are shown in Figure 2.6. The latitude distribution of this sample follows that of the BGPS as a whole, including peaking below $b = 0°$. The offset is related to the Sun's vertical displacement above the Galactic midplane (Schuller et al., 2009; Rosolowsky et al., 2010b). The only significant deviation is near $b = 0°$, where locally bright 8-μm emission along the midplane, excited by H II regions and OB stars, obscures more distant molecular cloud clumps. The BGPS 40″ flux density histograms (Fig. 2.6**b**) show that this sample contains, on average, brighter sources (median = 0.252 Jy) than the full Bolocat (median = 0.135 Jy). There are two likely origins of this bias. First, sources must have a $v_{\rm LSR}$ measurement from a dense-gas tracer; the HCO^+ detection fraction, in particular, is a strong function of BGPS flux density ($<$ 20% for $S_{1.1} < 0.1$ Jy; Shirley et al., 2013). Second, the selection criteria excluded sources with very low contrast or whose morphology does not correspond to dark regions in the GLIMPSE maps. Faint BGPS sources have low optical depth ($S_{1.1} = 0.1$ Jy corresponds to $\tau_8 \approx 0.07$), and would be difficult to distinguish against the variable Galactic 8-μm background.

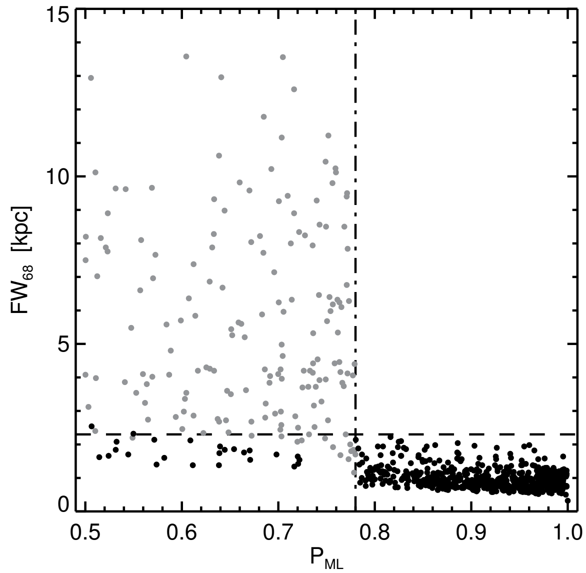

Figure 2.7 Comparison of the full-width of the 68.3% error bar (FW_{68}) against the integrated DPDF on the d_{ML} side of $d_{\tan}$ (P_{ML}). The vertical dot-dashed line represents the empirical cutoff at $P_{\mathrm{ML}} = 0.78$, and the horizontal dashed line marks $\mathrm{FW}_{68} = 2.3$ kpc. Objects shown in gray are below the P_{ML} cutoff and are more than 1 kpc from the tangent point. We defined $\mathrm{FW}_{68} \leq 2.3$ kpc as the criterion for a "well-constrained" distance estimate, which encompasses the objects in the bottom left corner, as well.

The distribution of measured mid-infrared source contrast is shown in Figure 2.6**c**, and has a median of 0.19. For images without the IRAC scattering correction, this value corresponds to an uncorrected median contrast of 0.15 (see Appendix 2.8.2), considerably lower than the minimum contrast ($C \approx 0.20$) used by Peretto & Fuller (2009) in their catalog of **Spitzer** IRDCs (which did not correct for IRAC scattering in the same manner). The majority of our sample consists of "low-contrast" sources that are missing from published catalogs of Galactic IRDCs.

2.5.2 Source KDA Resolutions

2.5.2.1 Distance Estimates and Constraints

We derived posterior DPDFs for the EMAF-selected BGPS sources by multiplying the kinematic distance DPDF by the two priors, and normalizing to unit total probability. By design, the prior DPDFs are broader than the peaks in $\mathrm{DPDF}_{\mathrm{kin}}$, so the resulting maximum-likelihood distances generally do not differ from the simple kinematic distances by more than ~ 0.1 kpc. To gauge the strength of the KDA resolution, two statistics were defined: the maximum-likelihood probability ($P_{\rm ML}$) as the integrated posterior DPDF on the $d_{\rm ML}$ side of the tangent point, and the full width of the 68.3% maximum-likelihood error bar ($\rm FW_{68}$). The ranges of these statistics are $0.5 \leq P_{\rm ML} \leq 1.0$ and $0.2\ {\rm kpc} \lesssim {\rm FW}_{68} \lesssim 15$ kpc, and a comparison between them for each object is shown in Figure 2.7. The nature of a double-peaked $\rm DPDF_{kin}$ leads to the sharp change in the distribution of $\rm FW_{68}$ near $P_{\rm ML} = 0.78$ (vertical dashed line). When the ratio of the peak probabilities of the kinematic distance peaks in the posterior DPDF becomes $\lesssim 3$, the error bars must include both to enclose sufficient probability. For objects with $P_{\rm ML} \geq 0.78$, the maximum-likelihood error bars enclose a single kinematic distance peak. We consider this set to have well-constrained distance estimates, and note that ${\rm FW}_{68} \leq 2.3$ kpc (horizontal dashed line). Objects with full-width error bars less than this value and are below the $P_{\rm ML}$ cutoff (lower-left corner of Fig. 2.7) are generally within ~ 1 kpc of the tangent distance. Because of the limited distance range available to these sources, their distance estimates should also be considered well-constrained. Combining these sets, we adopted ${\rm FW}_{68} \leq 2.3$ kpc as the criteria for well-constrained distance estimates.

Objects within a kiloparsec of $d_{\rm tan}$ have posterior DPDFs that are oftentimes asymmetric, and $d_{\rm ML}$ is not the best single-value representation of the distance. For these objects, we assigned them to the tangent distance group, and used $\overline{d}$ (and associated uncertainty) in the analysis that follows. A total of 618 sources in this sample have well-constrained distance estimates (80%). The KDA resolutions for these objects are recorded as "N" (near), "F" (far), or "T" (tangent distance) in column 8 of Table 2.3; $d_{\rm ML}$ is listed in column 10. The remaining objects are recorded as "U" (unconstrained) and have no distance estimate listed. For the tangent group, the weighted-average distance ($\overline{d}$) is listed in column 11, and is the preferred distance representation for these objects ($d_{\rm ML}$ is shown for comparison only). DPDFs for all sources are available in the BGPS archive.

2.5.2.2 Heliocentric Distances

For the remaining discussion, we consider only the 618 mid-infrared-dark BGPS sources whose

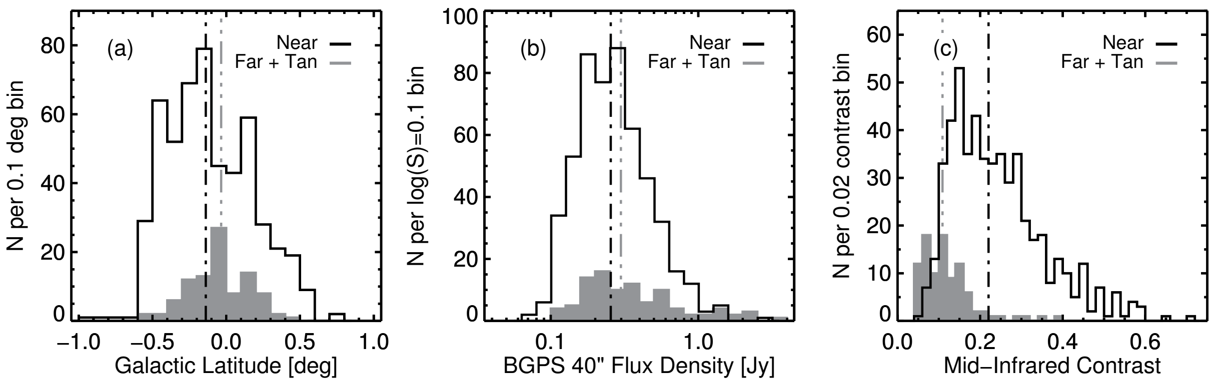

Figure 2.8 Comparison of source properties for objects with "near" vs. "far" KDA resolutions. Sources placed at the near distance are represented by open black histograms; tangent-point and far-distance sources are shown with filled gray histograms. Panels are as in Fig. 2.6. Vertical dot-dashed lines represent the median of each distribution.

distances are well-constrained (as defined above). Of this set, 70 were placed beyond the tangent point, with another 25 near d_{tan}, indicating the significant possibility of detecting molecular cloud clumps at the far kinematic distance using mid-infrared absorption. The comparisons of latitude distribution, BGPS 40″ flux densities, and mid-infrared contrast between the near and far subsets are shown in Figure 2.8. For the purposes of this discussion, objects at the tangent distance are grouped with those at the far kinematic distance. The latitude distributions (panel **a**) are very similar, with the near group being slightly wider, owing to the latitude-limiting effect of $\mathrm{DPDF}_{\mathrm{H_2}}$.

The histograms of BGPS 40″ flux densities (Fig. 2.8**b**) show that the far subset has a flatter distribution with a higher median than the near set. Since the source list for this study is mid-infrared-contrast limited, we do not expect to see low flux-density BGPS sources at the far distance; the low column density would not produce enough attenuation to be seen behind the significant foreground emission. The expected contrast as a function of heliocentric distance (represented by f_{fore}) may be computed by combining Equations (2.7), (2.9), and (2.11) into

$$C = (1 - f_{\mathrm{fore}})(1 - e^{-\Upsilon S_{1.1}}) \ , \tag{2.12}$$

where I_{emaf} and I_{MIR} from Equation (2.7) are equivalent to I_{min} and $\langle I_{\mathrm{MIR}} \rangle$ from Equation (2.11), respectively. A source with larger flux density may be at a farther $d_{\odot}$ and still meet the contrast

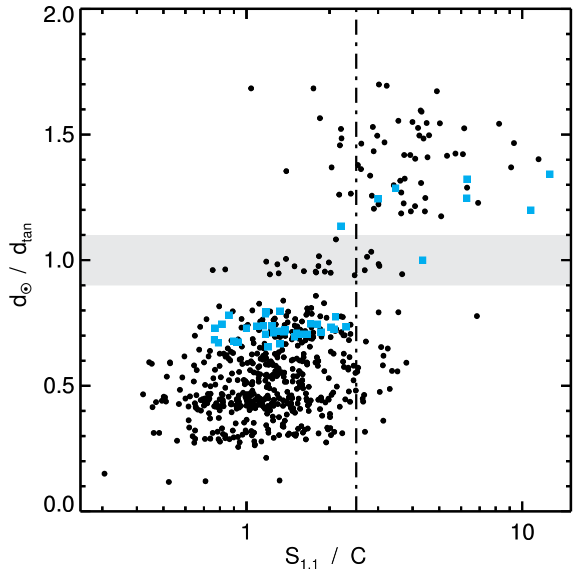

Figure 2.9 KDA resolution versus the ratio of BGPS 40″ flux density to mid-infrared contrast. The upper region represents the far kinematic distance, and the lower the near; the gray shaded region illustrates the band around $d_{\rm tan}$. Black dots mark the sources with well-constrained distance estimates; the subset of cyan squares are sources within W43 (see §2.6.1.1). The vertical dot-dashed line is drawn at $S_{1.1}/C = 2.5$. (A color version of this figure is available in the online journal.)

selection criterion.

Indeed, sources at the far kinematic distance have a lower median mid-infrared contrast ($C = 0.11$) than those at the near kinematic distance ($C = 0.22$), as shown in Figure 2.8**c**. Of the 313 objects with $C \geq 0.2$, only 12 (4%) are placed at the far kinematic distance, reinforcing the notion that dark mid-infrared absorption features must lie relatively nearby. For comparison, of the 63 sources with $C < 0.1$, 40 (63%) were placed at the far kinematic distance, indicating that the majority of EMAFs with very low contrast are at or beyond the tangent point.

An interesting empirical predictor of KDA resolution is shown in Figure 2.9. The ratio of resolved heliocentric distance over $d_{\rm tan}$ is plotted against the ratio of BGPS 40″ flux density over the mid-infrared contrast. For BGPS data, there appears to be a boundary at $S_{1.1}/C \approx 2.5$ that divides KDA resolutions. The exact value of this cutoff is dependent upon the (sub-)millimeter survey used, and there exists scatter across the boundary. It nevertheless suggests an additional means for KDA resolution when $\text{DPDF}_{\rm emaf}$ fails to return a well-constrained estimate.

Table 2.4. Maser Sources for Distance Comparison

Source Name	ℓ (°)	b (°)	$v_{\rm LSR}$ (km s^{-1})	Distance (kpc)	$N_{\rm BGPS}$ [a]	Ref.
G23.0-0.4	23.01	-0.41	81.5	$4.6^{+0.4}_{-0.3}$	6	1
G23.4-0.2	23.44	-0.18	97.6	$5.9^{+1.4}_{-0.9}$	2	1
G23.6-0.1	23.66	-0.13	82.6	$3.2^{+0.5}_{-0.4}$	2	2
W51 IRS2	49.49	-0.37	56.4	$5.1^{+2.9}_{-1.4}$	1	3
W51 Main[b]	49.49	-0.39	58.0	$5.4^{+0.3}_{-0.3}$	1	4

[a]Number of EMAF-selected BGPS sources within 15′ and 10 km s^{-1} of the maser location. See Fig. 2.13 for the comparison.

[b]H_2O maser; all others are CH_3OH masers

References. — (1) Brunthaler et al. (2009); (2) Bartkiewicz et al. (2008); (3) Xu et al. (2009); (4) Sato et al. (2010)

2.5.2.3 Galactocentric Positions

With well-constrained distance estimates, it is possible to construct a face-on view of the Milky Way. Sources with well-constrained KDA resolutions are plotted atop a reconstruction of the Milky Way from **Spitzer** data in Figure 2.10 (R. Hurt: NASA/JPL-Caltech/SSC) using either $d_{\rm ML}$ or $\overline{d}$ as described in §2.5.2.1. For clarity, the error bars, which account for small deviations from circular motion, are not shown. Some spiral structure is evident in the map of BGPS sources, notably portions of the Sagittarius Arm at $\ell \gtrsim 35°$, the Scutum-Centarus Arm / Molecular Ring at $\ell \lesssim 30°$, and the local arm / Orion spur within about a kiloparsec of the Sun (Churchwell et al., 2009). The kinematic restrictions on our sample led to the absence of objects within a ≈ 3.5 kpc radius of the Galactic center (dashed circle in the figure). Face-on views of the Galaxy derived from kinematic distances will not show narrow spiral features (like those in the background image) because of the local virial motions of individual molecular cloud clumps within larger complexes. Galactocentric positions are therefore "smeared-out" by approximately ± 0.4 kpc about the true kinematic distance for the complex as a whole. Each dot in the figure, however, should be thought of in terms of its DPDF, where the kinematic distance peaks have a FWHM of 1-2 kpc.

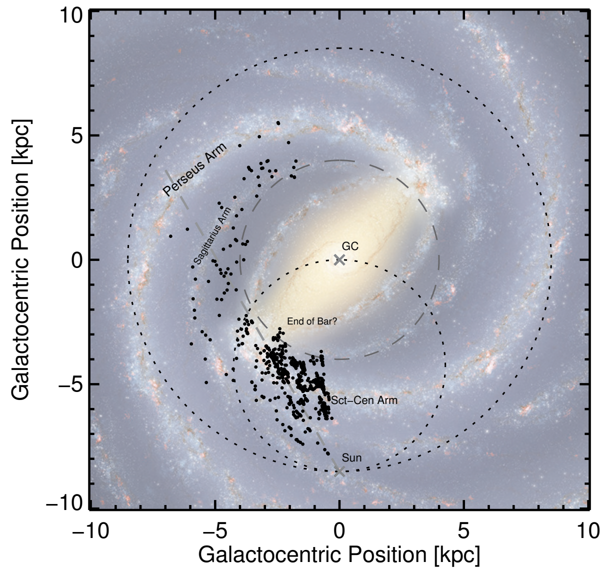

Figure 2.10 Face-on view of the Milky Way for sources with well-constrained KDA resolutions, plotted atop an artist's rendering of the Milky Way (R. Hurt: NASA/JPL-Caltech/SSC) viewed from the north Galactic pole. The image has been scaled to match the R_0 used for calculating kinematic distances. The outer dotted circle marks the Solar circle, and the inner dotted circle the tangent point as a function of longitude. The dashed circle at $R_{\rm gal} = 4$ kpc outlines the region influenced by the long Galactic bar (Benjamin et al., 2005), corresponding to the hashed regions in Fig. 2.5. The straight dashed gray line marks $\ell = 30°$ as a guide. Various suggested Galactic features are labeled. For clarity, distance error bars are not shown. (A color version of this figure is available in the online journal.)

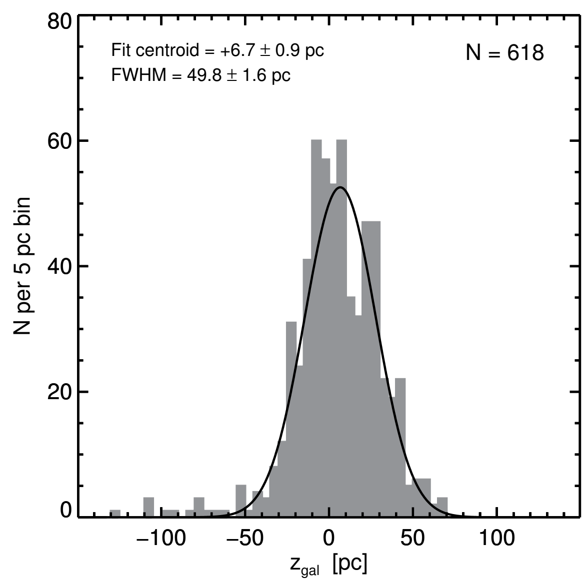

Figure 2.11 Vertical distribution of sources about the Galactic midplane. The filled gray histogram shows the distribution, while the black line represents a Gaussian fit to the histogram.

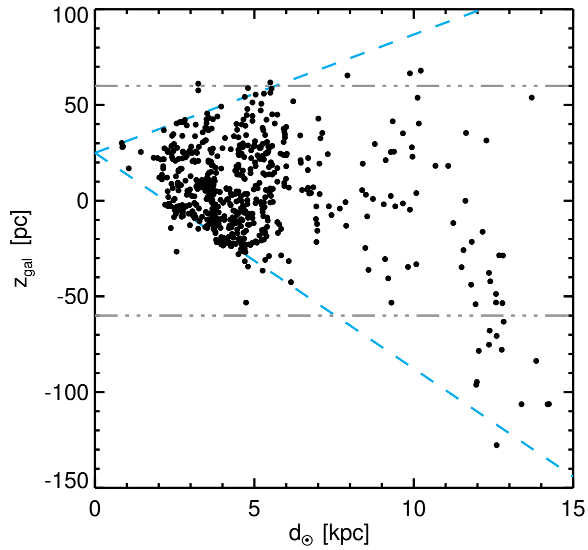

Figure 2.12 Derived vertical position of sources versus heliocentric distance. Diagonal cyan dashed lines represent the nominal $|b| = 0\overset{\circ}{.}5$ limit of the BGPS at $\ell = 30°$ for a vertical Solar offset above the Galactic midplane of 25 pc. Horizontal dot-dashed lines mark the FWHM of the ^{12}CO layer (Bronfman et al., 1988). (A color version of this figure is available in the online journal.)

KDA resolutions also allow the derivation of the vertical distribution of sources about the Galactic midplane. Vertical position is particularly affected by the KDA for higher-latitude sources ($|b| \gtrsim 0\overset{\circ}{.}4$). Calculation of vertical height (z) requires a proper accounting of the Sun's ≈ 25 pc vertical offset above the Galactic plane (Humphreys & Larsen, 1995; Jurić et al., 2008). The small scale height of Galactic molecular gas can lead to incorrect inferences about the vertical distribution of dense gas in the disk if z positions are calculated directly from Galactic coordinates without a correction for the Solar offset. The matrix needed to transform (ℓ, b, d) into $(R_{\mathrm{gal}}, \phi, z)$ is derived in Appendix 2.8.3.

The vertical distribution for the set of well-constrained BGPS sources is shown in Figure 2.11. A Gaussian fit to the distribution yields a half-width at half maximum of 25 pc, and a positive centroid offset of 7 pc. This scale height is approximately half that found by Bronfman et al. (1988) for ^{12}CO. This narrow result may, however, be a result of the limited Galactic latitude coverage of the BGPS. Analysis of the recent compact source catalog from the $\lambda = 870$ μm ATLASGAL survey (Contreras et al., 2013), which extends to $|b| = 1°$, shows that $\approx 20\%$ of their objects lie outside the BGPS latitude limits. To gauge the effect of limited latitude coverage, the derived z are plotted against heliocentric distance in Figure 2.12, with $|b| = 0\overset{\circ}{.}5$ shown for $\ell = 30°$ (the limits rotate to slightly more positive z for larger ℓ). For the region $d_\odot \leq 6$ kpc (which contains more than 80% of this sample), the BGPS does not fully probe the FWHM of the ^{12}CO distribution (dot-dashed lines). Other indicators of a larger scale height for star-formation regions include ^{13}CO clouds from the GRS (Roman-Duval et al., 2009), which have a FWHM $\approx$ 80 pc, and Galactic H II regions (FWHM $\approx$ 100 pc; Anderson et al., 2012).

2.5.3 Distance Comparisons with Other Studies

The quality of KDA resolutions for EMAF-selected BGPS sources was characterized through a comparison of distance estimates with values from the literature. In particular, comparison sets were chosen that used mostly orthogonal methodologies so that distance comparisons are largely free of correlated effects. The three sets described below are the use of maser parallax measurements, H I absorption features associated with molecular clouds, and near-infrared extinction measurements.

2.5.3.1 Maser Parallax Distances

Maser parallax measurements towards regions of high-mass star formation provide absolute distance validation comparisons. The Bar and Spiral Structure Legacy Survey (BeSSeL; Brunthaler et al., 2011) is conducting ongoing VLBI parallax measurements of CH_3OH and H_2O maser emission in star-forming regions across the Galactic plane. Such measurements provide very accurate distances out to $d_\odot \sim 10$ kpc, but the present overlap between published results and the BGPS is small (see Table 2.4 for the comparison set of maser sources used).

The comparison set was defined as objects from our sample whose angular separations and velocity differences were $\leq 15'$ and ≤ 10 km s^{-1}, respectively, from those of a published maser. These masers tend to be in regions of high-mass star formation, and such regions are on order $0\overset{\circ}{.}25$ in size. The velocity limits are related to the spread of virial velocities within such regions. A collection of 12 BGPS objects were associated with one of five masers; the distribution is noted in

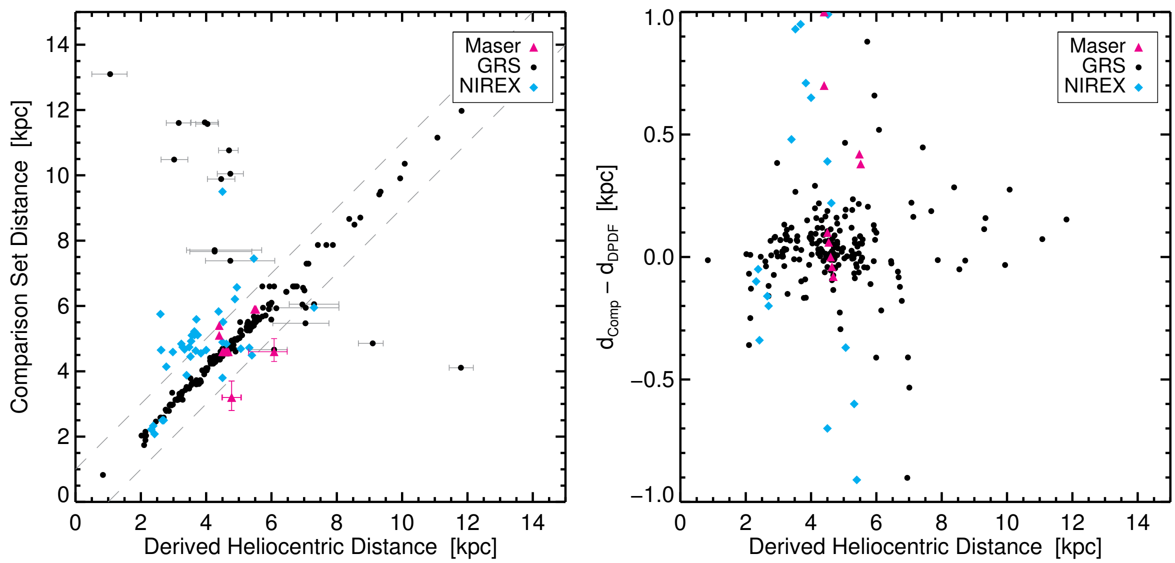

Figure 2.13 **Left:** Comparison of KDA resolutions derived from the DPDF with published distance estimates. Gray dashed lines represent ±1 kpc away from equality. Maximum-likelihood horizontal error bars are shown for GRS- and maser-associated sources lying outside this region. Vertical error bars for maser-associated sources come from Table 2.4. See text for discussion of systematic offset for NIREX sources. **Right:** A zoom-in on the region ±1 kpc from distance equality to better visualize the distance comparison for these objects. The vertical axis represents the comparison set distance minus the heliocentric distance from the DPDF. (A color version of this figure is available in the online journal.)

Table 2.4. To visualize the distance comparison, distances from Table 2.3 for each BGPS object are plotted against the measurements from the BeSSeL literature as magenta triangles in Figure 2.13. The gray dashed lines in the left panel represent ±1 kpc error margins, used for qualitative purposes. An object falling outside this region is said to have a "mismatching" distance estimate. For clarity in the figure, only mismatching objects have error bars shown; horizontal bars are from Table 2.3, and vertical bars are from the reference. The right panel is a zoom-in on the $\Delta d = \pm 1$ kpc region of the left panel (**i.e.**, within the dashed lines).

For the maser comparison set, only three sources fall outside the ±1 kpc region. Two BGPS objects (overlapping triangles in Fig. 2.13) are associated with the maser source G23.66-0.13, which has a parallax distance that disagrees with the derived (near) kinematic distance. Bartkiewicz et al. (2008) find that this object has a proper motion consistent with the parallax distance and the assumption of a flat rotation curve, but has a ≈ 35 km s^{-1} peculiar motion toward the Galactic center. This radial streaming motion makes its kinematic distance appear larger, and provides a cautionary example of the effects of non-circular motion on kinematic distance methods. The other mismatching source has a DPDF distance estimate skewed away from the simple kinematic distance due to the sharply-peaked DPDF$_{\rm emaf}$ caused by its bright millimeter flux density ($S_{1.1} =$ 3.9 Jy). The W51 region lies near the tangent point, so correct DPDF distance placement for these objects merely implies that the region's circular velocity is consistent with the rotation curve. The remaining EMAF-selected BGPS objects in this set have KDA resolutions that agree with the trigonometric parallax distance.

2.5.3.2 Galactic Ring Survey KDA Resolutions

For a larger distance comparison set, we used the KDA resolutions from the BU-FCRAO Galactic Ring Survey (GRS; Jackson et al., 2006). By matching ^{13}CO(1-0) emission morphology and spectra with H I absorption features, Roman-Duval et al. (2009) estimated the distances to some 750 molecular clouds in the inner Galaxy. Those authors used a combination of H I self-absorption (HISA)[7] and 21-cm continuum absorption features to positively resolve the KDA. These techniques exploit the spectroscopic dimension of H I surveys, where cold atomic hydrogen within dense molecular gas absorbs line emission from warm gas at the same $v_{\rm LSR}$ on the far side of the Galaxy or continuum radiation from H II regions. Distance resolutions from this method are subject to uncertainties from non-circular and radial streaming motions, but are directly comparable with the KDA resolutions of the DPDF method.

EMAF-selected BGPS objects were associated with cataloged ^{13}CO clouds based on spatial and kinematic proximity. The association volume was defined as a circle of radius $10'$ (approximately the median size of a GRS cloud; Roman-Duval et al., 2009), and a velocity spread equal to the ^{13}CO velocity dispersion, centered on the $(\ell, b, v_{\rm LSR})$ coordinates from the GRS catalog. A

[7] Absorption features caused by cold neutral hydrogen within molecular clouds are also called "narrow" self-absorption (HINSA) to distinguish them from the broader self-absorption features of diffuse H I clouds (**cf.** Li & Goldsmith, 2003).

total of 213 EMAFs lie within the association volume of one or more GRS clouds. To ensure the accuracy of the associated GRS KDA resolution, BGPS flux density maps were compared to both ^{13}CO intensity maps integrated over the velocity of the appropriate dense-gas tracer from Table 2.1, and H I 21-cm "on"−"off" integrated intensity (HISA) maps. For a handful of sources ($\sim 7\%$), a strong HISA signature was present within the BGPS source contour even though the associated GRS cloud was placed at the far kinematic distance. None of these objects was listed as having a 21-cm continuum source, so the absorption signature is the result of cold gas at the near distance. These discrepant objects may be the result of line-of-sight confusion, a slight velocity offset between the parent cloud and the BGPS object, or incorrect association with a ^{13}CO cloud. Whatever the cause, the KDA resolution for the associated GRS cloud was amended to "near" to reflect the HISA signature.

Roman-Duval et al. used the Clemens (1985) curve to derive heliocentric distances, and differences in rotation curve definition can cause distance-comparison discrepancies unrelated to the KDA (see Fig. 2.5). To eliminate potential systematic effects, the KDA resolution and v_{LSR} of each associated GRS cloud were mapped to a new heliocentric distance using the Reid et al. (2009b) rotation curve. In the comparison between the GRS-derived distance and those from the DPDFs (black dots in Fig. 2.13), nearly 92% of our distance resolutions match those of the GRS. This success rate is robust for the entire EMAF set, as enforcing a minimum mid-infrared contrast of $C \geq 0.15$ only increases the matching rate to 94%.

The 17 BGPS objects with mismatching distance resolutions are shown with horizontal error bars from the DPDFs. Those in the upper-left of Figure 2.13 have a large apparent mid-infrared absorbing column, but Roman-Duval et al. (2009) did not find evidence of self-absorbing H I. Conversely, those in the bottom-right have HISA signatures but were placed beyond the tangent point by the posterior DPDF. Examination by eye of this latter group showed that the two sources farthest from the one-to-one line have slight underestimates of I_{MIR} around the EMAF; the values in the postage-stamp image reflect dimmer nearby regions. The $\mathrm{DPDF}_{\mathrm{emaf}}$ in these cases selects the far kinematic distance peak despite the presence of HISA for these objects.

There are four objects whose GRS distance estimate is 5.5 kpc $\geq d_{\odot} \geq 8$ kpc and disagree with the DPDF-derived distance. These all lie within $\sim$ 1.5 kpc of the tangent point. Since the kinematic distance DPDFs do not have two fully distinct peaks in this region, the particular shape of $\mathrm{DPDF}_{\mathrm{emaf}}$ can have a significant impact on the derived single-distance estimators. The mismatches are due to the source being near d_{tan}, and not an incorrect KDA resolution. The remaining eleven mismatching sources in the upper-left of Figure 2.13 (GRS-far, $\mathrm{DPDF}_{\mathrm{emaf}}$-near) are moderately dark EMAFs ($0.1 \leq C \leq 0.3$) that show no signs of HISA at the velocity of the molecular cloud clump. About half of these lie at $|b| \geq 0\overset{\circ}{.}4$, and may not have enough H I backlighting at the far kinematic distance for a HISA signature to be visible; although Gibson et al. (2005) found self-absorption features out to more than $|b| = 2°$ in the Canadian (H I) Galactic Plane Survey. For sources in this quadrant of the figure, it is unclear which kinematic distance is correct. The future application of additional prior DPDFs may solve the small number of conflicting KDA resolutions, but the present method achieves very good correspondence with other distance estimates for molecular cloud clumps.

2.5.3.3 Near-Infrared Extinction Distances

Using a technique for measuring three-dimensional near-infrared Galactic extinction (NIREX; Marshall et al., 2006), Marshall et al. (2009) estimated the distances to over 1200 IRDCs identified by MSX in the inner Galactic plane (Simon et al., 2006a, hereafter S06). This approach compares the stellar colors of a section of sky with a Galactic stellar distribution model, and searches for sharp changes in color excess as a function of distance. Extinction measurements, like maser parallaxes, offer a kinematic-independent means of distance determination.

MSX dark clouds have a typical size of about an arcminute, so BGPS sources lying within 60″ of the centroid of a NIREX cloud were included in this comparison set. While there are about 275 objects from Marshall et al. (2009) within the spatial bounds of this study, only 38 EMAF-selected BGPS sources could be associated with a NIREX cloud. Peretto & Fuller (2009, hereafter PF09) noted that only a quarter of MSX IRDCs appear in their catalog of **Spitzer** dark clouds for a

Table 2.5. Effect of Dust Temperature on KDA Resolution

T_d	$N_{\rm wc}$ [a]	KDA Resolution[b]			GRS Comparison	
(K)		N	F	T	N	Rate[c]
15	625	416	175	34	218	77.5%
20	618	523	70	25	213	91.9%
25	605	547	33	25	198	91.9%

[a]Number of well-constrained distance estimates

[b]N = near; F = far; T = tangent point

[c]Distance matching success rate

variety of reasons. This selection effect, in combination with our requirement that an object be detected in one or more molecular line transitions, makes the number of matching clouds reasonable.

Objects from the NIREX comparison set are shown as cyan diamonds in Figure 2.13. Most of the points lie within 2 kpc of equality. Only one object has a wildly divergent distance estimate, G31.026-0.113 (BGPS #4653; $d_{\rm NIREX} \approx 9.5$ kpc), which has $C = 0.4$ and should have been detected with a strong near-infrared absorption signature at the near kinematic distance of $d_\odot = 4.5$ kpc. The collection of cyan diamonds with a systemic positive offset of 1.5 kpc warrants attention. There is a cluster of objects placed 2-4 kpc from the Sun. Most of these are at $\ell \lesssim 15°$, and uncertainties in both the rotation curve and stellar model in that region may be contributing to the offset. Mismatching NIREX distances beyond $d_\odot = 4$ kpc have divergent distance estimates of order the difference between the Clemens (1985) and Reid et al. (2009b) rotation curves for objects at that velocity.

2.6 Discussion

2.6.1 Kinematic Distance Discrimination

2.6.1.1 EMAFs as Distance Discriminators

The combination of millimeter-wave thermal dust emission observations with mid-infrared extinction is a powerful method for resolving the KDA for molecular cloud clumps. By starting from a catalog of (sub-)millimeter sources, this method is not limited to mid-infrared-identified IRDCs (catalogs of which often have large minimum contrast). We are therefore able to include low-column nearby sources as well as more distant objects. The 93% success rate compared to distances resolutions by the GRS team indicates that EMAFs can provide a powerful means for distance discrimination. Additionally, BGPS objects placed at the far kinematic distance that agree with the GRS distance indicate that EMAFs are visible beyond the tangent point with sufficient backlighting. In comparison with the HISA KDA-resolution technique employed by Roman-Duval et al. (2009), only 4% of BGPS objects were placed at the near kinematic distance, yet had no evidence of a HISA signature. Application of additional prior DPDFs may help to resolve these discrepancies.

The mid-infrared contrast distributions (Fig. 2.8**c**) of objects on either side of d_{tan} clearly show that objects placed at or beyond the tangent point have lower collective contrast, and would likely not be included in catalogs of IRDCs. These distributions are consistent with the notion that dark IRDCs ($C \gtrsim 0.2$) are nearby. Since the matching rate between DPDF-derived distances and those of the GRS is nearly independent of mid-infrared contrast, the present method extends robust KDA resolution to EMAFs with lower contrast, roughly doubling the number of molecular cloud clumps for which well-constrained distances may be derived.

In addition to improving upon the axiom "if IRDC then near" for KDA resolution, this method automatically accounts for the profile of the 8-μm intensity as a function of Galactic longitude (Fig. 2.2). The morphological matching process does not consider d_{tan}, and therefore offers a prior probability that is independent of the kinematic signature of a given object. The f_{fore} limit of visibility for a molecular cloud clump is simply a function of optical depth (Equation 2.12); for instance, an object with $S_{1.1} = 0.3$ Jy will have $C \geq 0.05$ for $f_{\mathrm{fore}} \leq 0.76$, but one with $S_{1.1} = 0.1$ Jy will not meet this contrast threshold if f_{fore} exceeds 0.33.

Heightened star-formation activity, which produces excess 8-μm emission in its immediate vicinity, does constitute a complicating factor in application of simple radiative transfer (Equation 2.6). These regions strain the assumption of smooth, axisymmetric Galactic emission. As an example, we analyzed the distance resolutions of objects in the W43 region. W43 is defined here by $31\overset{\circ}{.}5 \geq \ell \geq 29\overset{\circ}{.}5$, $-0\overset{\circ}{.}5 \geq b \geq 0\overset{\circ}{.}3$, and $80\ \mathrm{km\ s^{-1}} \geq v_{\mathrm{LSR}} \geq 110\ \mathrm{km\ s^{-1}}$ (as in Nguyen Luong et al., 2011), and is marked by a white box in Figure 2.5. Of the 43 EMAF-selected BGPS sources with well-constrained distance estimates in this region, 9 are placed at or beyond the tangent distance by the DPDF method. This is a slightly higher rate than the general sample, but is not significant. The W43 objects are plotted as cyan squares in Figure 2.9, and obey the empirical $S_{1.1}/C = 2.5$ limit for near versus far distance discrimination (§2.5.2.2). Only 8 objects could be associated with a GRS-identified ^{13}CO cloud (§2.5.3.2), and all but one have matching KDA resolutions; the outlier is one of the BGPS objects near the tangent point, where $\overline{d}$ is used, causing the > 1 kpc distance discrepancy. Although the observed Galactic plane consists of clumpy emission atop a more smooth Galactic emission pattern, the simple model used here still returns consistent KDA resolutions, even in more active regions.

While the KDA resolutions for EMAF-selected BGPS objects compare favorably with previously-published distance estimates, they are still based upon the assumption of circular orbits about the Galactic center. As highlighted by the case of CH_3OH maser G23.66-0.13, radial streaming motions can have a significant impact upon the derived kinematic distances. Anderson et al. (2012) presents a detailed analysis of uncertainties involved with the use of kinematic distances in the presence of non-circular motions. Future improvements in kinematic distance measurements will require a full three-dimensional vector model of Galactic motions.

Throughout this analysis we used $T_d = 20$ K as the temperature for converting BGPS flux densities into 8-μm optical depths. The effect of different assumed dust temperatures on KDA resolutions is not **a priori** predictable. Generating a new set of DPDF$_{\mathrm{emaf}}$ based on different temperatures, however, is straightforward. The results of KDA resolutions from the posterior DPDFs and GRS distance comparison statistics are shown in Table 2.5 for the range of T_d found by Battersby et al. (2011). A warmer dust temperature pushes more objects to the near kinematic distance to compensate for a smaller derived τ_8. Interestingly, the GRS success rate is unchanged

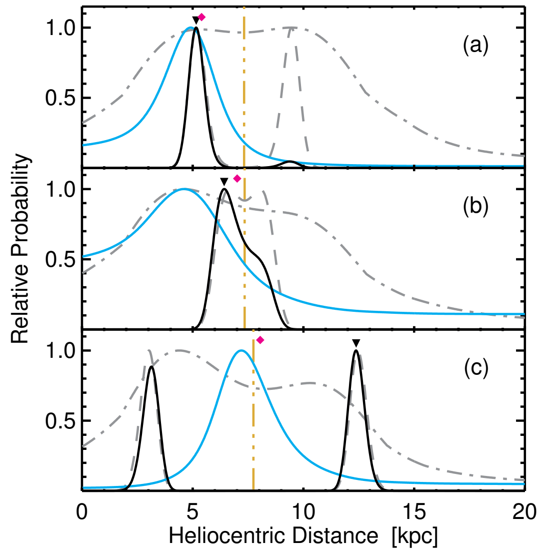

Figure 2.14 Example DPDFs for three possible cases. Shown in each panel are $\mathrm{DPDF}_{\mathrm{kin}}$ (dashed gray), $\mathrm{DPDF}_{\mathrm{emaf}}$ (solid cyan), $\mathrm{DPDF}_{\mathrm{H_2}}$ (dot-dashed gray), and posterior DPDF (solid black). The vertical dot-dashed line in each panel marks the tangent distance along that line of sight. The single-distance estimates are marked as black triangles (d_{ML}) and magenta diamonds ($\overline{d}$). Panel (**a**) represents a well-constrained KDA far from d_{tan}; (**b**) shows a source near d_{tan}; (**c**) illustrates a source with an unconstrained KDA resolution. (A color version of this figure is available in the online journal.)

for $T_d = 25$ K, but there are likely many unconsidered systematic effects at play.

The use of EMAFs as prior DPDFs for kinematic distance discrimination is directly applicable to all current and future (sub-)millimeter surveys of the Galactic plane. The advantage of starting with a sample of continuum-identified molecular cloud clumps is that mid-infrared extinction may be used to resolve the KDA for many objects that may not be dark enough to be included in IRDC catalogs (**e.g.**, S06, PF09).

2.6.1.2 Use of Different Distance Estimates

The distance probability density function (DPDF) formalism encodes all information about distance determinations for molecular cloud clumps, including likely distance and uncertainty. For some purposes, however, it is useful to have a single distance; §2.3.2 describes two possible options. In the use of the DPDFs produced here, it became apparent that different distance estimates were best applied to different situations. Examples of these situations are shown in Figure 2.14 to illustrate the difficulties encountered in extracting a single distance from a DPDF. For each panel, the black triangle and magenta diamond mark d_{ML} and $\overline{d}$, respectively.

The most common situation has well-separated kinematic distance peaks (**i.e.**, the molecular cloud clump is far from $d_{\tan}$) and the probability ratio of the two peaks is also large (**i.e.**, $P_{\mathrm{ML}} \geq 0.78$). The example BGPS #4484 (G030.629-00.029) is depicted in Fig. 2.14**a**. These objects fall under the "well-constrained" condition described in §2.5.2.1, whereby the maximum-likelihood error bars encompass only one kinematic distance peak. For this set of objects, d_{ML} is a reasonable collapse of the DPDF into a single value (with uncertainty). The second set of objects are those with kinematic distances within a kiloparsec of the tangent point. The kinematic distance DPDF for these objects have a shallow saddle feature at $d_{\tan}$, as seen in Fig. 2.14**b** for BGPS #4357 (G030.321+00.292). Since the main peak of the posterior DPDF is not symmetric, d_{ML} is not a robust reflection of the distance estimate. Therefore, we recommend using $\overline{d}$ for these sources (listed in Table 2.3 for this class of object) to more accurately reflect the available distance information. So long as the full-width of the error bars is less than 2.2 kpc (§2.5.2.1), these sources are considered to have well-constrained distance estimates. The final set of sources are those not meeting either of the above criteria, such as BGPS #3352 (G024.533-00.182; Fig. 2.14**c**). The kinematic distance options are well-separated, but the $\mathrm{DPDF}_{\mathrm{emaf}}$ does not place a strong discriminatory constraint. If it is desirable to use the distances for these sources, we recommend Monte-Carlo sampling distances from the full DPDF in order to include all distance information about the source (see §2.3.3).

Regardless of the method used, however, care should be taken to properly propagate the uncertainty in the distance placement. If only sources in the first category are used, it is important to remember that matching success rate with GRS-derived distances was $\approx 92\%$. This may be interpreted either as two sources in 25 were placed at the wrong distance or that there is a 92% confidence in each of the distance placements.

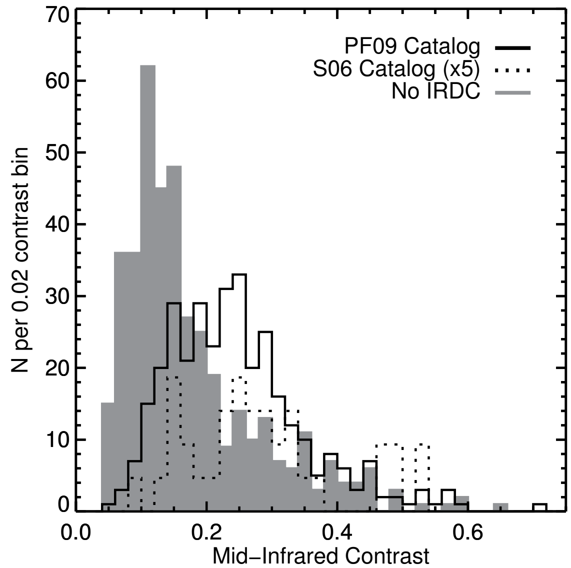

Figure 2.15 Histograms of EMAF contrast as a function of IRDC identification. The black open histogram depicts objects associated with the PF09 catalog; black dotted shows those associated with S06 ($\times 5$ for clarity); filled gray represents objects associated with neither catalog.

2.6.2 EMAF vs. IRDC

In this paper, we introduce the nomenclature "Eight-Micron Absorption Feature" (EMAF) for dust-continuum-identified molecular cloud clump whose emission morphology matches an absorption feature in mid-infrared maps of the Galactic plane. Many of these objects are quite dark ($C \gtrsim 0.2$) and are identified in IRDC catalogs. To better understand the overlap between the EMAF and IRDC designations, we searched through both the S06 and PF09 catalogs to find the closest IRDC to the location of peak BGPS flux density. A total of 361 (46%) EMAF-selected BGPS sources lay within the semi-major axis distance of the centroid of a cloud from one or both of the catalogs.

By definition, IRDCs have a large mid-infrared contrast, but EMAFs are selected from thermal dust emission catalogs. As such, the two groups have different contrast distributions, as seen in the histograms of Figure 2.15. BGPS sources that are associated with an object in the PF09 or S06 catalogs are plotted as black solid and dotted lines, respectively. The filled gray histogram represents EMAFs not associated with any IRDC. The bulk of the non-IRDC objects have $C \lesssim 0.2$, once again confirming that the EMAF designation allows the use mid-infrared observations for KDA resolution for objects with low contrast. Of particular interest are the higher-contrast ($C \gtrsim 0.2$) BGPS sources which do not appear in either IRDC catalog (filled gray). One striking example is the source G035.478-00.298 (BGPS #5631), shown in the lower-right corner of Figure 2.3**a**. These objects suggest that it is easier to identify molecular cloud clumps from (sub-)millimeter data than to try to find intensity decrements in $\lambda = 8\ \mu$m images.

The EMAF-derived KDA resolutions of the IRDC objects place only a small fraction at the far kinematic distance (3% and 7% for S06 and PF09, respectively). The measured EMAF contrast for such objects is generally $C \lesssim 0.2$, reinforcing the notion that dark IRDCs are nearby. Since the fractions of IRDCs placed beyond $d_{\tan}$ are comparable to the GRS distance mismatch rate (§2.5.3.2), these subsets are not significant.

2.6.3 Implications for Galactic Structure

2.6.3.1 Galactic 8-μm Emission

A quick glance at the GLIMPSE mosaics suggests that the Galactic distribution of 8-μm light cannot be described simply by a smooth, diffuse model. Distances derived using the assumption of smooth emission, however, compare favorably to those derived from 21-cm H I absorption (GRS; Roman-Duval et al., 2009) and near-infrared extinction mapping (NIREX; Marshall et al., 2009). These results suggest that Galactic 8-μm emission may be primarily composed of a diffuse component punctuated by regions of active star formation. The R12 model of mid- and far-infrared emission is a greatly simplified reflection of the Galaxy, neglecting to account for individual small-scale features. Yet its match, spatially and spectrally, to existing Galactic plane observations indicates its power. The model, therefore, provides a solid basis for our distance estimation technique based on using a broad distance prior to distinguish between kinematic distance peaks.

While the R12 model was constructed to broadly match the multi-band observations of the Milky Way, it is useful to also compare it with external galaxies. **Spitzer** observations of SINGS

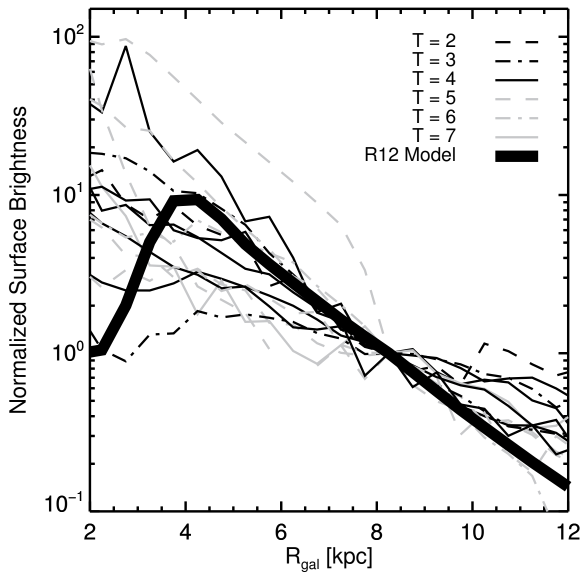

Figure 2.16 Azimuthally-averaged surface brightness (λ = 8 μm) as a function of galactocentric radius. A sample of 17 SINGS galaxies is plotted in various line styles according to Hubble stage (T), and are normalized to $R_{\mathrm{gal}} = 8.5$ kpc. The thick black line depicts the surface brightness of the mid-infrared emission model (Fig. 2.1). The depression in the model profile at low R_{gal} is due to the dust hole in the R12 model (§2.4.1.3).

galaxies (Kennicutt et al., 2003) yield a surface density of in-band emission. Figure 2.16 depicts the IRAC Band 4 surface brightness, normalized to $R_{\mathrm{gal}} = 8.5$ kpc for a collection of 17 SINGS galaxies, with galaxies represented by their Hubble stage $(T)^8$. To enable comparison, the R12 model was viewed externally (see Fig. 2.1) and emission was annularly integrated, following the SINGS analysis; plotted as a thick black line. The slope of the Milky Way model at $R_{\mathrm{gal}} \gtrsim 4$ kpc is comparable to the ensemble, but the drop in emission due to the dust hole carved out by R12 near the Galactic center does not seem to match extragalactic observations. The model was designed to match observations from **within** the disk from the Sun's location, however, so discrepancies in the integrated 8-μm emission profile in the inner $R_{\mathrm{gal}} \leq 3$ kpc are likely not relevant.

2.6.3.2 Spiral Structure

Recent large (sub-)millimeter Galactic plane surveys are making it possible to trace the spiral

[8] Hubble stage is a continuous numerical representation of the Hubble type for a galaxy. $T = 0$ corresponds to an S0 galaxy, and the Milky Way (SBb-c) is $T \approx 4$ (Binney & Merrifield, 1998, p.155).

structure of the disk. Comparing the well-constrained KDA resolutions of BGPS sources with an artist's conception of the Galaxy based on **Spitzer** data (Fig. 2.10), glimmers of organization begin to appear. While the regions that trigger the collapse of molecular cloud clumps are likely very localized along spiral density waves, features in the face-on map of the Galaxy derived from kinematic distances are quite smeared out. Even small (≈ 5 km s^{-1}) peculiar motions can lead to ≈ 0.4 kpc variations in heliocentric position, making it difficult to precisely trace the locations of spiral features.

Two major features suggest themselves from the data in Figure 2.10: one near and one far. First, the nearby collection of sources at $\ell \lesssim 30°$ seem to form part of a round feature that extends into the fourth quadrant. This feature has been identified as both the Molecular Ring (**cf.** Dame et al., 2001) and the Scutum-Centarus Arm (**cf.** Dobbs & Burkert, 2012). It is not possible to distinguish between these postulates in the northern plane; careful distance determinations for southern sources is required. Mapping the exact location of where this feature meets the long Galactic bar also depends on choice of rotation curve (the Clemens, 1985 curve places this collection of sources at the tangent distance, whereas the Reid et al., 2009b curve does not), and may also be influenced by non-circular motions. Parallax measurements to masers in this region will help establish a benchmark for the long Galactic bar, including position angle with respect to the Sun $-$ Galactic Center line.

The other feature to note in Figure 2.10 is the Sagittarius Arm beyond the tangent circle (smaller dotted circle). These molecular cloud clumps are visible at the far kinematic distance because of backlighting provided by the Perseus Arm. By the same token, BGPS sources in the Perseus Arm are not visible as EMAFs due to both the large amount of 8-μm light in the foreground and the lack of any significant backlighting source. We note a collection of sources around $\ell \sim 30°$ which appear at the far kinematic distance. In the underlying image, there is a void between the Sagittarius Arm and the long Galactic bar. The two possibilities, therefore, are the existence of an arm structure at that location or that those sources were improperly assigned the far kinematic distance. The analysis of the W43 region, however, suggests that active regions do not significantly misplace molecular cloud clumps at the far kinematic distance, so distinguishing between the possibilities is unclear.

2.7 Summary

We developed DPDFs as a new method for distance determinations to molecular cloud clumps in the Galactic plane. Starting from a kinematic distance derived from molecular line observations as the likelihood, prior DPDFs may be applied in a Bayesian manner to resolve the KDA. In this study, we used two external data sets as priors: mid-infrared absorption features, and the Galactic distribution of molecular gas.

The dust in molecular cloud clumps detected by (sub-)millimeter Galactic plane surveys should absorb mid-infrared light and be visible against the broad diffuse PAH emission near $\lambda = 8\ \mu$m. Starting from the BGPS catalog of dust-continuum-identified molecular cloud clumps, we identified 770 EMAFs in the **Spitzer**/GLIMPSE mosaics. EMAFs may be thought of as generalized IRDCs, and are characterized by their selection from (sub-)millimeter data. With this collection of objects, simple radiative transfer arguments, and a model of Galactic mid-infrared stellar and dust emission, we developed a morphological matching scheme to compare dust emission and absorption. When using the GLIMPSE mosaics to measure apparent absorption features, it is imperative to account for scattering of light within the IRAC camera. This scattering, in concert with the instrumental calibration method, means that diffuse emission will appear brighter than it really is; bright emission will tend to fill in absorption features. The scattered light changes the apparent contrast of absorption features, in addition to any derived properties (such as optical depth, mass, etc.).

Well-constrained KDA resolutions were obtained for 618 objects in this sample: 527 at the near kinematic distance, 31 at the tangent point, and 70 at the far kinematic distance. To corroborate our distance discriminations, we used VLBI maser parallax measurements, KDA resolutions from the GRS, and near-infrared extinction distances as comparison sets. Of the 13 objects associated with maser parallax measurements, none had a discrepant KDA resolution. Distance comparisons with the GRS yielded a 93% success rate nearly independent of mid-infrared contrast. Comparison with the NIREX distances showed only one discrepant KDA resolution, with the remainder being within identified systematic effects. These comparisons illustrate the validity of the present method, including the placement of some EMAFs at the far kinematic distance (approximately 12 distance-atched GRS sources are beyond $d_{\rm tan}$).

Approximately half of the set of EMAFs are associated with an object from the IRDC catalogs of S06 and PF09. Objects associated with IRDCs are mostly relatively dark ($C \gtrsim 0.2$); the remainder being largely low-contrast ($C \lesssim 0.2$). Interestingly, there are a handful of moderately dark ($C \gtrsim 0.3$) EMAFs that are absent from these IRDC catalogs. This suggests that it is perhaps easier to identify molecular cloud clumps first in (sub-)millimeter data, then investigate their mid-infrared properties.

KDA resolutions for EMAFs from the BGPS catalog reveal hints of Galactic structure. Foremost, most detectable Galactic molecular cloud clumps are in the Molecular Ring / Scutum-Centarus Arm feature between the Sun and the Galactic center. The Sagittarius Arm outside $\ell = 30°$ is suggested by a collection of EMAFs beyond the tangent point, visible due to backlighting from the more-distant Perseus Arm.

The derivation of DPDFs allows for probabilistic determination of distances to molecular cloud clumps across the Galactic plane. By introducing the concept of an EMAF, we were able to use the mid-infrared GLIMPSE data to resolve the KDA for many more sources than is possible with extant catalogs of IRDCs. Although this method applies only to $\sim 10\%$ of the BGPS catalog, the DPDF framework allows for the incorporation of additional prior DPDFs to expand the number of molecular cloud clumps with well-constrained distance estimates.

2.8 Appendices

2.8.1 Computing a Band-Averaged Dust Opacity: Spitzer IRAC Band 4

The apparent dust opacity of extinction features in broadband images is related to the dust opacity as a function of frequency and the spectrum of the light being absorbed. The IRAC Band 4 bandpass includes several distinct emission features from PAH molecules, as well as the complex behavior of the dust opacity near the 10-μm silicate feature. In this appendix, we derive a band-averaged dust opacity $\langle\kappa\rangle_{\rm band}$ for use with the GLIMPSE mosaics.

Given the simple radiative transfer model of Equation (2.6), the intensity transmitted through a dust cloud is

$$I_{\rm trans} = I_{\rm back}\ e^{-\langle\tau\rangle_{\rm band}}\ . \tag{2.13}$$

where $I_{\rm back}$ is the background light (from the cloud to large heliocentric distance), and $I_{\rm trans}$ is the transmitted light exiting the cloud on the near side (**i.e.**, not including emission between the cloud and the observer). The band-averaged optical depth is related to the apparent dust opacity by $\langle\tau\rangle_{\rm band} = \Sigma\ \langle\kappa\rangle_{\rm band}$, where Σ is the mass surface density of dust. The band-averaged dust opacity is therefore

$$\langle\kappa\rangle_{\rm band} = -\frac{1}{\Sigma}\ln\left(\frac{I_{\rm trans}}{I_{\rm back}}\right)\ . \tag{2.14}$$

The ratio of intensities is computed from the band average over each quantity,

$$\begin{aligned}\frac{I_{\rm trans}}{I_{\rm back}} &= \frac{\langle I_{{\rm back},\nu}\ e^{-\tau_\nu}\rangle_{\rm band}}{\langle I_{{\rm back},\nu}\rangle_{\rm band}} \\ &= \frac{\int R_{\rm band}(\nu)\ I_{{\rm back},\nu}\ e^{-\tau_\nu}\ d\nu}{\int R_{\rm band}(\nu)\ I_{{\rm back},\nu}\ d\nu}\ ,\end{aligned} \tag{2.15}$$

where the ν subscript denotes that quantity as a function of frequency, and $R_{\rm band}(\nu)$ is the relative frequency response per unit power for the instrument. Since both averages are over the same response bandpass, the usual normalization terms cancel.

The typical radius of an IRDC is small ($\sim$ 1 pc; Rathborne et al., 2006) compared to the accumulated path length (D) for the diffuse background (several kiloparsecs), so the intensity $I_{{\rm back},\nu} = \int j_\nu\ ds$ may be approximated by $I_{{\rm back},\nu} = j_\nu\ D$, where j_ν is the emission coefficient.

Additionally, the relative response per unit power, $R_{\text{band}}(\nu)$, is proportional to $(1/h\nu)\ S_{\text{band}}(\nu)$, where $S_{\text{band}}(\nu)$ is the relative response per photon[9] . Canceling frequency-independent quantities and inserting the intensity ratio into Equation (2.14) yields the desired band-averaged dust opacity,

$$\langle\kappa\rangle_{\text{band}} = -\ln\left[\frac{\int \nu^{-1}\ S_{\text{band}}(\nu)\ j_\nu\ e^{-\kappa_\nu}\ d\nu}{\int \nu^{-1}\ S_{\text{band}}(\nu)\ j_\nu\ d\nu}\right]\ . \tag{2.16}$$

For the emission spectra (j_ν), we used the dust emission models of Draine & Li (2007), which contain a mixture of grain sizes in addition to a variable PAH mass fraction (q_{PAH}). The various emission spectra were derived by irradiating the dust with starlight intensity fields having a tunable minimum value U_{min} relative to the local interstellar radiation field ($U = 1$).

The choice of dust opacity model (κ_ν) has a nontrivial effect on the derived band-averaged opacity. Three different models were analyzed, and are shown in Table 2.6. First is the OH5 model (dust grains with thin ice mantles, coagulating at 10^6 cm^{-3} for 10^5 years; Ossenkopf & Henning, 1994) used for determining the dust opacity at $\lambda = 1.1$ mm for the BGPS. The remaining models were presented in Weingartner & Draine (2001, hereafter WD01), with updated normalizations given by Draine (2003)[10] . The second of the three models is the $R_V = 3.1$ Milky Way model, tried even though this value of the color excess per magnitude extinction is consistent with the diffuse ISM and does not hold for regions of dense gas. The final model utilizes case A for $R_V = 5.5$ (consistent with observations of molecular clouds), which sought to minimize the extinction differences between observation and model, while also including a penalty term to keep dust grain volume from exceeding abundance/depletion limits (WD01). For each dust model, $\langle\kappa\rangle_{\text{band}}$ was computed for three values each of q_{PAH} and U_{min} (Table 2.6).

The minimum value of the starlight intensity field has very little effect on the band-averaged dust opacity, meaning that the derived $\langle\kappa\rangle_{\text{band}}$ are valid for a wide range of environments. An order of magnitude change in the assumed PAH mass fraction causes only a 2% difference in the derived opacity for the OH5 model, but the spread is 10% for the others. Draine & Li (2007) cite $q_{\text{PAH}} = 4.58\%$ as best matching observations of the Milky Way.

[9] Obtained from `http://irsa.ipac.caltech.edu/data/SPITZER/docs/irac/calibrationfiles/spectralresponse`.
[10] `http://www.astro.princeton.edu/~draine/dust/dustmix.html`

Table 2.6. Computed $\langle\kappa\rangle_{\rm band}$ for IRAC Band 4

Dust Emission[a]		$\langle\kappa\rangle_{\rm band}$		
$q_{\rm PAH}$ (%)	$U_{\rm min}$	OH5[b]	WD01-3.1[c] ($\mathrm{cm}^2\ \mathrm{g}^{-1}$)	WD01-5.5[d]
0.47......	0.1	1175	858	909
	1.0	1177	861	911
	10.0	1178	864	915
2.50......	0.1	1157	796	843
	1.0	1158	800	847
	10.0	1160	804	852
4.58......	0.1	1150	777	823
	1.0	1151	779	825
	10.0	1152	782	828

[a]Dust emission model from Draine & Li (2007)

[b]Extinction from Ossenkopf & Henning (1994, Table 1, Column 5)

[c]R_V = 3.1 extinction from WD01, with updated normalizations from Draine (2003)

[d]R_V = 5.5, Case A, extinction from WD01, with updated normalizations from Draine (2003)

The OH5 model is the preferred description of dust at (sub-)millimeter wavelengths (**cf.** Rathborne et al., 2006; Schuller et al., 2009; Aguirre et al., 2011), and has been used by previous studies for estimating $\langle\kappa\rangle_{\rm band}$ for IRAC Band 4 images (Butler & Tan, 2009; Battersby et al., 2010). That model was computed from theory for coagulated grains (aggregates of smaller particles, with some voids) surrounded by an ice mantle. In contrast, the WD01 dust models utilized simple geometry (PAH molecules for very small grains, and graphite and olivine spheres for larger grains) and sought to fit a dust size distribution to parameterized observed extinction.

The connection to observed extinction in the infrared led us to choose the WD01 $R_V = 5.5$ model for this work. Following Draine & Li (2007), we used $q_{\rm PAH} = 4.58\%$ and $U_{\rm min} = 1.0$ to compute $\kappa_8 = \langle\kappa\rangle_{\rm band} = 825$ cm^2 g^{-1} of dust. For comparison, the corresponding value from the OH5 model yields $\kappa_8 = 1167$ cm^2 g^{-1}, a $\approx 40\%$ difference. We note that the preferred WD01 model predicts a BGPS dust opacity $\kappa_{1.1} = 0.272$ cm^2 g^{-1} of dust, approximately one quarter the value from OH5.

2.8.2 Spitzer IRAC Scattering Correction Factors

The IRAC camera on the **Spitzer** Space Telescope suffers from internal scattering within the detector arrays, particularly Bands 3 and 4. The scattering is such that a fraction of the incident light on a pixel is distributed throughout the entire array (Reach et al., 2005; IRAC Instrument Handbook). Image frames are converted into physical units (MJy sr^{-1}) using point-source calibration data; point-source aperture photometry is therefore accurate because the calibration takes into account the light scattered out of the aperture and into blank sky pixels. Observed extended emission, however, has light from other areas of the array scattered **into** each pixel as well, and so will appear brighter than it really is given the point-source calibration. For the broad, diffuse emission of the Galactic plane in IRAC Band 4, there is a roughly constant positive offset of the measured intensity in each frame.

For absorption features in the Band 4 images (EMAFs), however, the scattering cannot be corrected for by simple multiplicative aperture corrections. Because bright emission from surrounding regions is scattered into an EMAF, it will have a lower apparent contrast. To correct for this effect, an estimate of the scattered light in a frame must be **subtracted** from each frame (S. Carey 2010, private communication), as was done in this study for a pixel-by-pixel comparison between GLIMPSE and synthetic 8-μm images.

Quantities such as contrast and optical depth for EMAFs may be derived from the GLIMPSE mosaics (**e.g.**, Butler & Tan, 2009; PF09), but a scattering correction must be applied (**cf.** Battersby et al., 2010). Because careful subtraction of scattered light is not always necessary for a given application, correction factors may be derived for quantities measured directly from the IRAC Band 4 data. In this appendix, we derive correction factors for mid-infrared contrast (C), 8-μm optical depth (τ_8), and foreground fraction of 8-μm emission ($f_{\rm fore}$).

For regions of broad diffuse emission punctuated by dark clouds, the observed intensities I_0 and I_1 of the background and EMAF, respectively, are related to the actual intensities S_0 and S_1 by

$$I_0 = S_0 + X \ , \text{ and} \tag{2.17}$$

$$I_1 = S_1 + X \ , \tag{2.18}$$

where $X = \xi\ S_0$ is the amount of scattered light, approximated by the fraction $\xi = (1 - 0.737) = 0.263$ of incident diffuse light scattered throughout the array, and 0.737 is the infinite-aperture correction for Band 4 from Reach et al. (2005). Rearranging to compute the true intensities from observed quantities yields

$$S_0 = \frac{I_0}{(1+\xi)} \ , \text{ and} \tag{2.19}$$

$$S_1 = I_1 - I_0 \frac{\xi}{(1+\xi)} \ . \tag{2.20}$$

While the subtractive correction (Equation 2.17) for the diffuse background is equivalent to a multiplicative correction (Equation 2.19), correcting for the intensity within the EMAF is more complicated. To compute the true contrast of an EMAF, we begin with Equation (2.11):

$$C_{\rm true} = 1 - \frac{S_1}{S_0} \ . \tag{2.21}$$

Inserting Equations (2.19) and (2.20), the true contrast becomes

$$\begin{aligned} C_{\rm true} &= \left(1 - \frac{I_1}{I_0}\right)(1+\xi) \\ &= C_{\rm meas}\,(1+\xi)\ , \end{aligned} \tag{2.22}$$

where $C_{\rm meas}$ is the quantity measured directly from the GLIMPSE images. The measured contrast will be **smaller** than reality, leading to an underestimation of optical depth and other quantities.

With a measured EMAF contrast, it is possible to estimate the optical depth of a cloud or the foreground fraction of diffuse emission, given an assumption about the other. Equation (2.12) may be rearranged to solve for either quantity in terms of the other. The optical depth (from which follows surface mass density) is given by

$$\begin{aligned} \tau_{8,\rm true} &= -\ln\left[1 - \frac{C_{\rm true}}{1 - f_{\rm fore}}\right] \\ &= -\ln\left[1 - \frac{C_{\rm meas}\,(1+\xi)}{1 - f_{\rm fore}}\right]\ , \end{aligned} \tag{2.23}$$

assuming a model that yields $f_{\rm fore}$ (as in Butler & Tan, 2009). The true value of τ_8 will be larger by up to a factor of two for $C_{\rm meas} \lesssim 0.5$. If, instead, the foreground fraction is desired given an external estimate of τ_8 (such as from (sub-)millimeter thermal dust continuum data; as in PF09), the true value is given by

$$\begin{aligned} f_{\rm fore,true} &= 1 - \frac{C_{\rm meas}\,(1+\xi)}{1 - e^{-\tau_8}} \\ &= \left(1 - \frac{C_{\rm meas}}{1 - e^{-\tau_8}}\right)(1+\xi) - \xi \\ &= f_{\rm fore,meas}\,(1+\xi) - \xi\ . \end{aligned} \tag{2.24}$$

The actual foreground fraction will be smaller than that measured directly from GLIMPSE images, with the difference becoming less at large $f_{\rm fore}$. Any $f_{\rm fore,meas} \lesssim 0.2$ maps to zero true foreground fraction, as negative values are not physical; such values arise from uncertainty in C and the derivation of τ_8 from (sub-)millimeter data.

2.8.3 The Vertical Solar Offset and Converting $(\ell, b, d_\odot)$ into $(R_{\rm gal}, \phi, z)$

Deriving Galactocentric positions of objects in the Milky Way requires a coordinate transformation of the triad $(\ell, b, d_\odot)$, where $d_\odot$ is the heliocentric distance along the line of sight toward (ℓ, b). The Galactic coordinate system was defined assuming the Sun is at the midplane of the disk (Blaauw et al., 1960), but more recent studies have measured a vertical solar offset of $\approx$ 25 pc above the midplane (Humphreys & Larsen, 1995; Jurić et al., 2008). Since the vertical scale height of the molecular gas layer in the disk is small (HWHM $\approx$ 60 pc; Bronfman et al., 1988), neglecting to account for the solar offset may introduce a systematic bias in the derived vertical distributions of components of the Galactic disk.

The coordinate transformation is done in cartesian coordinates. First the triad $(\ell, b, d_\odot)$ is converted into local cartesian coordinates, where the x-axis is directed along the Sun $-$ Galactic Center line, and z points north out of the plane,

$$\begin{pmatrix} x_1 \\ y_1 \\ z_1 \end{pmatrix} = \begin{pmatrix} d_\odot \cos l \cos b \\ d_\odot \sin l \cos b \\ d_\odot \sin b \end{pmatrix} . \tag{2.25}$$

The local coordinates are transformed to the Galactocentric frame by (1) rotation by 180° in the $x-y$ plane to place the $+x$-axis pointing away from the GC, (2) translation of the coordinate axes to place the origin at the GC, and finally (3) rotation by the angle θ in the $x-z$ plane to place the $+x$-axis along the Galactic midplane (rather than along the Sun $-$ Galactic Center line). The transformation may be written as

$$\begin{pmatrix} x_{\rm gal} \\ y_{\rm gal} \\ z_{\rm gal} \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 1 & 0 & 0 \\ \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & R_0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ 1 \end{pmatrix} , \tag{2.26}$$

where the lateral translation requires an augmented (affine translation) matrix. The resulting

Galactocentric cartesian coordinates are

$$\begin{pmatrix} x_{\rm gal} \\ y_{\rm gal} \\ z_{\rm gal} \end{pmatrix} = \begin{pmatrix} R_0 \ \cos\theta - d_\odot \ (\cos l \ \cos b \ \cos\theta + \sin b \ \sin\theta) \\ -d_\odot \ \sin l \ \cos b \\ R_0 \ \sin\theta - d_\odot \ (\cos l \ \cos b \ \sin\theta - \sin b \ \cos\theta) \end{pmatrix} . \tag{2.27}$$

The rotation angle $\theta = \sin^{-1}(z_0/R_0)$, where $z_0 = 25$ pc, corrects for the Sun's vertical displacement above the midplane. Galactocentric positions in the cylindrical coordinates $(R_{\rm gal}, \phi, z)$ may be extracted from Eqn. (2.27) in the usual manner. The rotation by θ is most important for the derived $z_{\rm gal}$, and has negligible effect on $R_{\rm gal}$ and ϕ.

Chapter 3

Distance Catalog Expansion Using Kinematic Isolation of Dense Molecular Cloud Structures With ^{13}CO(1-0)

We present an expanded distance catalog for 1,802 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 two new methods for use with our Bayesian distance probability density function (DPDF) methodology. The dense-gas tracers (**e.g.**, HCO^+(3-2), NH_3(1,1)) used to derive line-of-sight velocities for kinematic distances have a $\sim 50-85\%$ detection rate for BGPS catalog objects. To expand the set of objects with velocity information, we utilize the Galactic Ring Survey ^{13}CO(1-0) data, but use of the brightest emission peak along a line-of-sight yields a velocity in agreement with HCO^+(3-2) only $\sim 84\%$ of the time. To improve this rate to $\approx 94\%$, we present a method for morphologically extracting velocities from the ^{13}CO data using millimeter emission as prior information. The outline of a BGPS source is used to select a region of the GRS ^{13}CO data, along with a reference region to subtract enveloping diffuse emission, to produce a line profile of ^{13}CO matched to the BGPS source. We also present 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. We demonstrate this prior using catalogs of masers with trigonometric parallaxes and H II regions with robust KDA resolutions. The resulting distance catalog contains well-constrained distance estimates for 21% of the BGPS V2 catalog, with typical distance uncertainties $\lesssim 0.5$ kpc. Approximately 75% of the well-constrained sources lie closer than 6 kpc from the Sun, concentrated in the Scutum-Centarus

arm. Galactocentric positions of objects additionally trace out portions of the Sagittarius, Perseus, and Outer arms in the first and second Galactic quadrants.

3.1 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., 2010b) 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 molecular cloud structures detected by continuum surveys. The primary input to this Bayesian method is a kinematic distance, whereby a measured line-of-sight velocity ($v_{\rm LSR}$) is combined with a Galactic rotation curve. Geometric considerations reveal that two heliocentric distances ($d_\odot$) may correspond to a single $v_{\rm LSR}$ 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}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.**, HCO^+(3-2), NH_3(1,1), CS(2-1); Evans, 1999) have been targeted by pointed observations in an effort to derive single-component $v_{\rm LSR}$ for these structures (**cf.** 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 $v_{\rm LSR}$.

The largest single Galactic molecular line data set to date is the compiled ^{12}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 ^{13}CO is $\sim 50 - 90\times$ optically thinner (depending on Galactocentric location) and is a better tracer of structure within GMCs. In large-scale ^{13}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 ^{13}CO(1-0) excitation also significantly increases the detection of multiple cloud structures along a given line of sight. Direct $v_{\rm LSR}$ comparison of the brightest ^{13}CO peak along the line of sight towards continuum-detected objects with the $v_{\rm LSR}$ from a dense-gas tracer like HCO^+(3-2) yields a $\sim 81\%$ agreement rate (Shirley et al., 2013). In this work we present a technique for extracting a ^{13}CO

v_{LSR} for molecular cloud structures using the morphology and flux density from continuum data as prior information to increase the agreement rate to $\sim 94\%$.

The kinematic distance likelihoods from ^{13}CO data expand the catalog of objects with detected v_{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_3OH and H_2O masers from the Bar and Spiral Structure Legacy (BeSSeL; Brunthaler et al., 2011) Survey and Japanese VERA project[1] 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 3.2 describes the data used and the distance-computation code. An overview of the DPDF formalism and relevant Galactic kinematics is given

[1] `http://veraserver.mtk.nao.ac.jp`

in Section 3.3. The estimation of molecular cloud clump v_{LSR} from ^{13}CO data is described in Section 3.4, while the new prior DPDFs are outlined in Section 3.5. Section 3.6 presents the results, and implications of this work are discussed in Section 3.7. A summary is presented in Section 3.8.

3.2 Data

3.2.1 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^2 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 $|b| \leq 0\overset{\circ}{.}5$, 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, 8,594 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.[2]

The BGPS data pipeline removes atmospheric signal using a principle component analysis technique that discards common-mode time-stream 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] Available through IPAC at
`http://irsa.ipac.caltech.edu/data/BOLOCAM_GPS`

3.2.2 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[3] $n_{\rm eff} \gtrsim 10^3$ cm^{-3}. All 6,194 BGPS V1 sources with $\ell \geq 7\overset{\circ}{.}5$ were simultaneously observed in the J=3−2 rotational transitions of HCO^+ (ν = 267.6 GHz) and N_2H^+ (ν = 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^+ 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 $29° \leq \ell \leq 31°$ were observed in the J=2−1 rotational transition of CS (ν = 97.98 GHz) using the Arizona Radio Observatory 12m telescope on Kitt Peak (Y. Shirley, 2012, private communication). The lower $n_{\rm eff}$ 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 November 2013 at the HHT using the J=2−1 rotational transition of $C^{18}O$ (ν = 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_3 near 24 GHz with the Robert F. Byrd Green Bank Telescope. The (1,1) inversion is the strongest NH_3 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_3(1,1) range from 61% in the Gemini OB1 molecular cloud to 72% in the inner Galaxy.

[3] 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).

3.2.3 The Galactic Ring Survey

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

3.2.4 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`),[4] and is publicly available.[5] 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.

3.3 DPDFs and Galactic Kinematics

3.3.1 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

[4] `http://www.exelisvis.com/ProductsServices/IDL.aspx`
[5] `https://github.com/BGPS/distance-omnibus`

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

$$\mathrm{DPDF}(d_\odot) = \mathcal{L}(v_\mathrm{LSR}, l, b; d_\odot) \prod_i P_i(l, b; d_\odot) \ , \tag{3.1}$$

where $\mathcal{L}(v_\mathrm{LSR}, l, b; d_\odot)$ is the kinematic distance likelihood function,[6] 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_LSR of the product of the projected rotation curve along (ℓ, b) as a $f(d_\odot, v_\mathrm{LSR})$ with the molecular line emission spectrum (itself a function of v_LSR), and is generally double-peaked for Galactocentric positions within the solar circle (**i.e.**, $R_\mathrm{gal} < R_0$). 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_ML), and the weighted-average first-moment distance ($\overline{d}$). The d_ML is simply the distance corresponding to the largest probability in the posterior DPDF, while $\overline{d}$ is computed via $\overline{d} = \int_0^\infty d_\odot \ \mathrm{DPDF} \ \mathrm{d}(d_\odot)$. If the DPDF is well-constrained to have a single peak, d_ML and $\overline{d}$ will be nearly equivalent, but in cases where the KDA resolution is not well-constrained, these distance estimates may be substantially different and $\overline{d}$ becomes a biased estimator of the distance while d_ML becomes less reliable. For well-constrained distances away from the tangent point we recommend d_ML, but $\overline{d}$ represents a better distance estimate for objects near the tangent point due to the complex, multimodal structure of $\mathcal{L}(v_\mathrm{LSR}, l, b; d_\odot)$ in these regions.

[6] In EB13, this was referred to as $\mathrm{DPDF}_\mathrm{kin}$. We shift to the present notation to distinguish between the kinematic distance likelihood function and the prior DPDFs used to resolve the KDA.

Table 3.1. Galactic Rotation Curve Parameters

Parameter	Reid et al. (2009b)	Used in EB13[a]	Reid et al. (2014) Model A5	Used Here
R_0 (kpc)	8.4 ± 0.6	8.51	8.34 ± 0.16	8.34
Θ_0 (km s^{-1})	254 ± 16	244	240 ± 8	240
$\frac{d\Theta}{dR}$ (km s^{-1} kpc^{-1})	0	0	-0.2 ± 0.4	0
$U_\odot$ (km s^{-1})	10.0	11.1	10.7 ± 1.8	10.7
$V_\odot$ (km s^{-1})	5.2	12.24	15.6 ± 6.8	15.6
$W_\odot$ (km s^{-1})	7.2	7.25	8.9 ± 0.9	8.9
$\overline{U_s}$ (km s^{-1})	0	0	2.9 ± 2.1	0
$\overline{V_s}$ (km s^{-1})	-15	-6	-1.6 ± 6.8	0

[a]Values from M. Reid (2011, private communication), Schönrich et al. (2010)

3.3.2 Galactic Kinematics

3.3.2.1 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 6-dimensional trigonometric parallax data (position, velocity) to fit Galactic rotation curve models (**cf.** Reid et al., 2009b). Unlike curves based on tangent-point data that require assumed values for R_0 and Θ_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 $\Theta(R) = \Theta_0 + \frac{d\Theta}{dR}(R - R_0)$ 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 $(R_0, \Theta_0, \frac{d\Theta}{dR})$, the values of the solar peculiar motion $(U_\odot, V_\odot, W_\odot)$, and the mean deviant (non-circular) motions of high-mass star forming regions (HMSFRs) towards the Galactic center $(\overline{U_s})$ and in the direction of Galactic rotation $(\overline{V_s})$.

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.[7] The resulting fit parameters are listed in Table 3.1, along with the corresponding values used here and in previous works.

For simplicity, we apply a strictly flat rotation curve, which is consistent with the measured $\frac{d\Theta}{dR} = -0.2 \pm 0.4$ km s^{-1} kpc^{-1}. Additionally, we do not apply a correction for the counterrotation of HMSFRs $(\overline{V_s})$, also consistent with the values from model A5. While the nonzero mean source motion towards the Galactic center $(\overline{U_s})$ from model A5 is marginally significant, omitting this correction for molecular cloud clumps only results in a distance error of $\lesssim 0.2$ kpc for the vast majority of BGPS V2 sources. Contours in Figure 3.1 (**left**) illustrate the projected v_{LSR} derived from model A5 as a function of Galactocentric position for the northern Galactic plane.

3.3.2.2 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_{LSR} all map to a narrow set of heliocentric distances. Regions of the Galaxy that violate these assumptions must be excluded

[7] 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 (**cf.** Reid et al., 2009b; McMillan & Binney, 2010; Schönrich et al., 2010; Honma et al., 2012). While both $\Theta_o + V_\odot$ and $V_\odot - \overline{V_s}$ are highly constrained, the individual parameters are highly correlated (Reid et al., 2014).

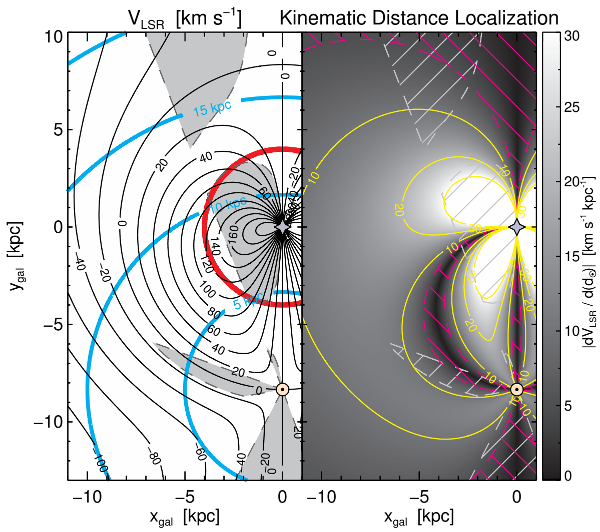

Figure 3.1 Galactic kinematics for distance determination. **Left**: Contours mark lines of constant projected $v_{\rm LSR}$ based on the (flat) rotation curve of model A5 from Reid et al. (2014). Regions for which kinematic distances are not computed (kinematic avoidance zones) are shown in gray with 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 $R_{\rm gal} = 4$ kpc, inside which the Galactic bar dominates the kinematics. **Right**: The magnitude of the derivative of $v_{\rm LSR}$ with respect to $d_\odot$ across the Galactic plane (grayscale). Magenta hashing identifies regions where $|dv_{\rm LSR}/d(d_\odot)| \leq 5$ km s^{-1} kpc^{-1} (**i.e.**, a projected peculiar motion of 5 km s^{-1} will yield a ≥ 1 kpc distance error). Kinematic avoidance zones are marked by gray hashing, and yellow contours mark constant values of $|dv_{\rm LSR}/d(d_\odot)|$.

from kinematic distance calculations, as derived distances may be off by a factor of two or more (**cf.** Reid et al., 2009b). We label these regions "kinematic avoidance zones."

Toward the center of the Milky Way, the long Galactic bar at $R_{\rm gal} \lesssim 3 - 4$ kpc (Fux, 1999; Rodriguez-Fernandez & Combes, 2008; Reid et al., 2014) introduces strong radial 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 $|\ell| \lesssim 20°$. To quantify this exclusion, EB13 (their Figure 5) defined two regions in the longitude-velocity diagram for which $\mathcal{L}(v_{\rm LSR}, l, b; d_\odot)$ is not computed. For $7.^\circ5 \leq \ell \leq 21°$,[8] the upper exclusion region is bounded by $v_{\rm LSR} = (3.33\ \text{km s}^{-1})\times\ell(°) + 15\ \text{km s}^{-1}$, and includes the higher-velocity gas inside the bar. The lower region excludes the 3-kpc expanding arm, and is bounded by $v_{\rm LSR} = (2.22\ \text{km s}^{-1})\times\ell(°) -$

[8] The lower longitude limit on this region is the lowest ℓ 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.

16.7 km s^{-1}. The effective regions in Galactocentric coordinates for these zones, computed for the rotation curve described above, are shown in Figure 3.1 (**left**) as gray areas towards the Galactic center. The allowed velocities in this longitude region correspond to the Scutum-Centarus arm feature (**cf.** Dame & Thaddeus, 2011).

Toward the Galactic anti-center ($\ell = 180°$), orbital circular motion is nearly perpendicular to the line of sight, making the projection of $v_{\rm LSR}$ onto the rotation curve highly subject to small peculiar motions of gas. We therefore define the region $\ell = 180° \pm 20°$ as an additional kinematic avoidance zone (see Figure 3.1), 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.**, $\ell = 90°$ and 270°) present fairly flat $v_{\rm LSR}(d_\odot)$ 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 §3.5.2). The flatness of the projected rotation curve (**i.e.**, $dv_{\rm LSR}/d(d_\odot) \leq 5$ km s^{-1} kpc^{-1}; magenta hashing in Figure 3.1 (**right**)) is not symmetric about $\ell = 90°$ and 270°, but instead follows the tangent circle toward the Galactic center. The expected virial motion within a HMSFR is ≈ 5 km s^{-1}, and distance errors of $\gtrsim 1$ kpc due to virial motions are not desirable. At $\ell \lesssim 70°$, this region around the tangent point becomes narrower and therefore acceptable. The rotation curve derivative is steep enough for $\ell \gtrsim 100°$ that the kinematic avoidance zone should only be defined over $70° \leq \ell \leq 100°$. In this range, however, the small $dv_{\rm LSR}/d(d_\odot)$ occurs only at small heliocentric distance (and consequently $v_{\rm LSR} \sim 0$ km s^{-1}). Examination of the $\ell - v$ diagram reveals that the Cygnus X region ($\ell \approx 80°$) extends from about -15 km s^{-1} $\leq v_{\rm LSR} \leq 20$ km s^{-1} (falling in the avoidance zone), with the better-defined Perseus and Outer arms visible at more negative $v_{\rm LSR}$. To fully encompass the Cygnus X region as a kinematic avoidance zone, we therefore limit the kinematic avoidance zone here to $v_{\rm LSR} \geq -15$ km s^{-1}, and the resulting region in Galactocentric coordinates is visible in Figure 3.1. The corresponding zone near $\ell = 270°$ would need to be defined based on gas kinematics near the Carina tangent.

3.4 Molecular Cloud Clump ^{13}CO Velocity Extraction

3.4.1 Comparison of GRS ^{13}CO Spectra with Dense Gas Tracers

The most straightforward method for using CO data to assign $v_{\rm LSR}$ to molecular cloud clumps is to choose the brightest emission peak along the line of sight. This has been used by various groups (**cf.** Russeil et al., 2011; Eden et al., 2012) in lieu of observations of molecular line transitions with higher $n_{\rm eff}$ 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 $v_{\rm LSR}$ assignment.

o illustrate the challenges of simply choosing the brightest ^{13}CO peak for BGPS objects, Figure 3.2 compares that velocity with available $\mathrm{HCO^+}$(3-2) (S13) and $\mathrm{NH_3}$(1,1) (Dunham et al., 2011b) observations. The extracted $v_{\rm LSR}$ from ^{13}CO agrees with the dense-gas value to within 5 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}CO spectrum; S13 show that using the single GRS spectrum at the location of peak $\lambda = 1.1$ 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 relatively little CO emission on the far side of the Galaxy beyond the solar circle behind this feature in the first quadrant, the GRS did not observe negative $v_{\rm LSR}$. 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. Several such objects are illustrated in Figure 3.2 as pink dots with $\mathrm{HCO^+}$(3-2) $v_{\rm LSR} < -5$ km s^{-1}. Of the 1,846 BGPS objects with detected dense-gas $v_{\rm LSR}$ in the GRS overlap region, however, only 18 are beyond the solar circle.

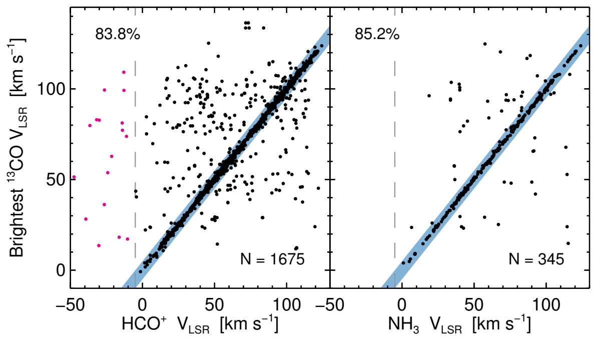

Figure 3.2 Comparison of the $v_{\rm LSR}$ of the brightest ^{13}CO emission feature from the GRS data towards a BGPS object compared to the (**left**) HCO^+(3-2) velocities of Shirley et al. (2013) and (**right**) NH_3(1,1) velocities of Dunham et al. (2011b). The vertical gray dashed line in each panel represents the lower $v_{\rm LSR}$ bound of GRS observations. Objects depicted in pink ($v_{\rm LSR} < 5$ km s^{-1}) would, therefore, not be detectable in the GRS data. The light blue region marks the $|\Delta v_{\rm LSR}| \leq 5$ km s^{-1} used to compute the matching statistic shown at the top of each panel.

3.4.2 Morphological Spectrum Extraction Technique

3.4.2.1 General Description

While the $\sim 85\%$ velocity matching rate for the brightest ^{13}CO feature along a line of sight is encouraging, the low $n_{\rm eff}$ of ^{13}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}CO emission from a particular molecular cloud structure from its more diffuse envelope (Rosolowsky et al., 2010a). The extracted spectrum is computed as $T_{\rm source}(v) = T_{\rm on}(v) - T_{\rm off}(v)$, where $T_{\rm on}(v)$ is the average of the GRS spectra at the location of BGPS pixels within the Bolocat label masks (see Rosolowsky et al., 2010b), weighted by BGPS flux density. Each BGPS pixel is assigned the ^{13}CO spectrum of the nearest (ℓ, b) GRS

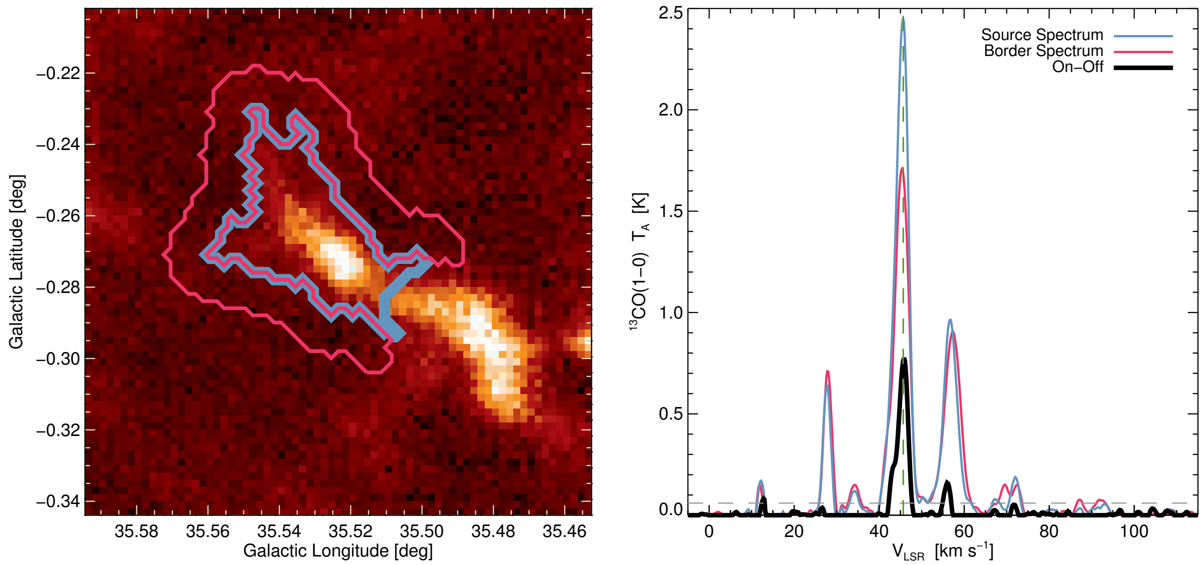

Figure 3.3 Demonstration of the morphological spectrum extraction technique for example source G035.523-00.273. **Left**: The image depicts the BGPS V2 flux density on a linear scale from -0.13 to 0.60 Jy beam^{-1}. The blue contour is the Bolocat source outline, as described in Rosolowsky et al. (2010b), and the pink contour depicts the off-source region, one GRS resolution element in width, excluding pixels associated with the continuum source to the lower right. **Right**: The filtered ^{13}CO(1-0) spectra for this object. Blue and pink spectra represent the $T_{\rm on}(v)$ and $T_{\rm off}(v)$, respectively, while black represents $T_{\rm source}(v)$, used to estimate the $v_{\rm LSR}$ of the denser gas associated with the molecular cloud clump.

pixel; the same ^{13}CO spectrum ($22''$ pixels) may therefore be assigned to more than one BGPS pixel ($7\farcs2$). 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 $T_{\rm off}(v)$ contributions only from lines of sight penetrating the more diffuse CO envelope. The $46''$ width of the off region corresponds to ~ 1.1 pc at $d_\odot = 5$ 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 $T_{\rm off}(v)$ may be computed. For these sources, $T_{\rm on}(v)$ is used in place of $T_{\rm source}(v)$, 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 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 H_2O Southern Galactic Plane Survey (HOPS; Walsh et al., 2011) observed several molecular transition lines around 24 GHz, mapping nearly 700 molecular cloud clumps in NH_3(1,1) (Purcell et al., 2012). Comparing the maximum solid angle the BGPS recovers ($19.6'^2$; $\theta_{\rm max} \approx 300''$; 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 NH_3(1,1) emission brighter than ≈ 0.2 K (the average root-mean-squared noise temperature of the HOPS NH_3(1,1) spectra; Purcell et al., 2012), and is appropriate for using as a prior for extracting ^{13}CO spectra for molecular cloud clumps.

The averaged ^{13}CO spectra were filtered using a kernel constructed to match the typical widths of ^{13}CO features (FWHM $\approx 2 - 5$ 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 $T_{\rm source}(v)$ with peak intensity less than the nominal GRS noise level. The resulting spectrum is clipped to contain only positive (emission) values. Figure 3.3 illustrates the morphological spectrum extraction process for object G035.523-00.273. The "Off" region in the left panel (pink outline) excludes BGPS pixels associated with the source to the lower right, and the averaged GRS spectrum from this region is shown in pink in the right panel. At most velocities, the background spectrum (blue) closely approximates $T_{\rm on}(v)$, with the notable exception near $v_{\rm LSR} =$ 45 km s^{-1}.

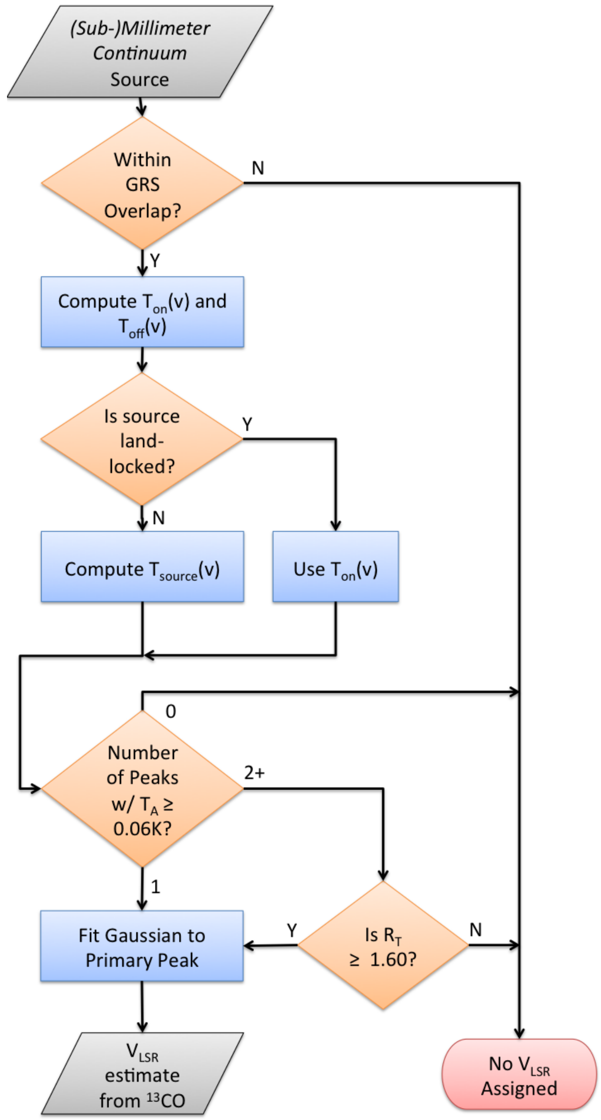

Figure 3.4 Flow chart for the assignment of $v_{\rm LSR}$ from the morphologically extracted ^{13}CO spectrum for a given molecular cloud structure.

3.4.2.2 Extracting $v_{\rm LSR}$ from $T_{\rm source}(v)$

This new technique effectively yields a pointed catalog of "position-switched" ^{13}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 $v_{\rm LSR}$ for kinematic distance computation. Two parameters control the quality and quantity of extracted spectra: the minimum T_A threshold for peaks in $T_{\rm source}(v)$ ($T_{A,\rm thresh}$),

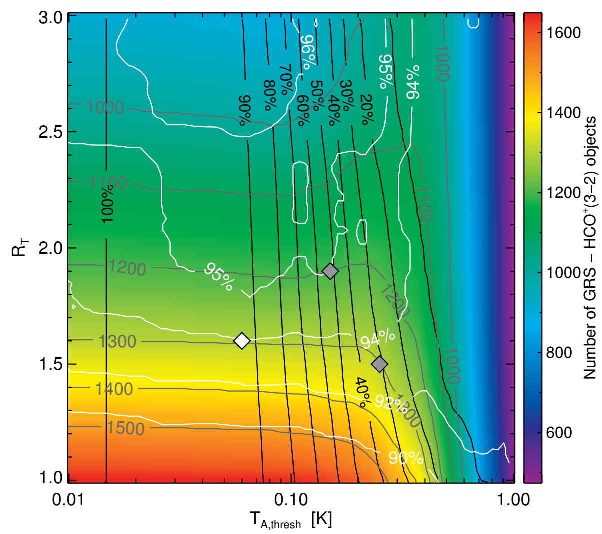

Figure 3.5 Optimization of the $T_{A,\mathrm{thresh}}$ for detection and R_T limit for multiply peaked spectra. Background color scale and gray contours denote the number of BGPS V2 objects that possess both HCO^+(3-2) and morphologically extracted ^{13}CO velocities (N_{comp}). White contours mark the fraction of N_{comp} whose v_{LSR} from the two sources agree to better than 5 km s^{-1}. Black contours show the fraction of N_{comp} whose morphologically extracted spectrum contains more than one peak above $T_{A,\mathrm{thresh}}$. The white diamond indicates the chosen values of $T_{A,\mathrm{thresh}}$ and R_T while the gray diamonds are discarded choices (see text).

and the degree to which a single v_{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 3.3, the primary peak is near $v_{\mathrm{LSR}} = 45$ km s^{-1}, and a secondary peak near $v_{\mathrm{LSR}} = 57$ km s^{-1}, with $R_T \approx 2.5$.

The algorithm for v_{LSR} extraction from $T_{\mathrm{source}}(v)$ is illustrated in the flow chart of Figure 3.4. For objects within the GRS coverage region, $T_{\mathrm{source}}(v)$ is examined for any peak above $T_{A,\mathrm{thresh}}$. Next, the number of contiguous regions above $T_{A,\mathrm{thresh}}$ is counted; for two or more independent peaks, R_T is computed. There are three points in the process where a null v_{LSR} may be returned: (1) the molecular cloud structure is outside the limits of the GRS, (2) there are no peaks in $T_{\mathrm{source}}(v)$ above $T_{A,\mathrm{thresh}}$, and (3) for multiply-peaked spectra there is no clearly dominant peak (**i.e.**, $R_T <$ limit). For the remaining multiply-peaked and all single-peaked spectra, a Gaussian is fit to the primary peak.

To determine the optimal values of these parameters for v_{LSR} extraction, we computed the following as functions of the two tunable parameters: (1) the number of objects ($N_{\rm comp}$) having both a valid ^{13}CO $v_{\rm LSR}$ and HCO$^+$ $v_{\rm LSR}$ from S13, (2) the fraction of $N_{\rm comp}$ whose ^{13}CO and HCO$^+$ velocities agree to ≤ 5 km s^{-1} (the approximate virial motion within HMSFRs), and (3) the fraction of $N_{\rm comp}$ whose ^{13}CO spectra contain multiple peaks above $T_{A,\rm thresh}$. Based on the topography of the parameter space, as shown in Figure 3.5, we chose values of $T_{A,\rm thresh} = 0.06$ K, and $R_T = 1.60$ (white diamond) to best balance the number of BGPS V2 objects in the GRS - HCO$^+$ overlap with the fraction of those sources whose $v_{\rm LSR}$ agree to better than 5 km s^{-1}.

We note that at the selected values for $T_{A,\rm thresh}$ and R_T, over 90% of the morphologically extracted spectra are multiply-peaked (black contours), yet the velocity-agreement fraction is virtually unchanged down to $\sim 40\%$ multiplicity. Two discarded values are shown as gray diamonds for comparison. The diamond at $(T_{A,\rm thresh}, R_T) = (0.25$ K, $1.5)$ has a threshold at approximately twice the rms noise of the GRS data cubes, and follows the same $N_{\rm comp} = 1300$ contour the white diamond is on. The matching rate at this point, however, is 93%. A point with a higher matching percentage (95%) is the gray diamond at $(T_{A,\rm thresh}, R_T) = (0.15$ K, $1.9)$. While this is one percent better than the white diamond, there are ≈ 100 fewer objects for which a morphological $v_{\rm LSR}$ may be extracted from the ^{13}CO data. The white diamond was ultimately a choice made to balance $N_{\rm comp}$ with the matching percentage.

3.4.2.3 Comparison with HCO$^+$(3-2)

With the optimized parameter values for $T_{A,\rm thresh}$ and R_T, there are 1,325 objects which possess both an HCO$^+$(3-2) detection and a valid morphologically extracted spectrum from the GRS ^{13}CO data. The velocity comparison is shown in Figure 3.6, which echoes Figure 3.2, except that the correspondence has grown to $\approx 94\%$. As with Figure 3.2, pink circles at the left of the plot mark continuum-detected molecular cloud structures beyond the solar circle ($R_{\rm gal} > R_0$) that also have detectable ^{13}CO emission in the foreground. It is likely these objects would have the correct $v_{\rm LSR}$ assigned (to 94% confidence) if the GRS had extended velocity coverage to negative $v_{\rm LSR}$. The histogram of the velocity difference for all points in Figure 3.6 is well-fit by a Gaussian

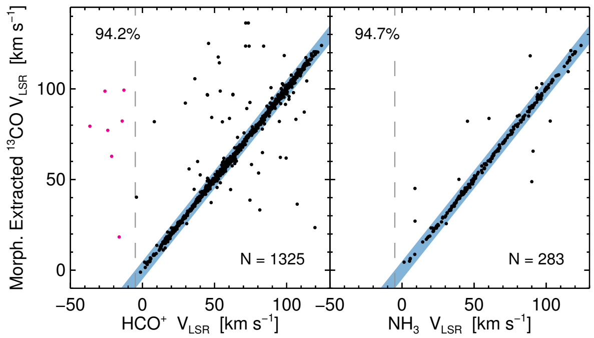

Figure 3.6 Morphologically extracted ^{13}CO spectrum $v_{\rm LSR}$ comparison with dense-gas tracers. Colors and lines as in Figure 3.2. **Left**: Velocity comparison for the $N = 1,325$ BGPS V2 objects with a valid morphologically extracted ^{13}CO spectrum and HCO$^+$(3-2) $v_{\rm LSR}$ from Shirley et al. (2013). **Right:** Velocity comparison for the $N = 283$ BGPS V2 objects with a valid morphologically extracted ^{13}CO spectrum and NH_3(1,1) $v_{\rm LSR}$ from Dunham et al. (2011b).

with FWHM ≈ 1.1 km s^{-1}, the channel width of the HCO$^+$ spectra (S13).

3.4.3 The ^{13}CO Kinematic Distance Likelihood Function

Using the optimized parameters for computing a morphologically extracted spectrum from the GRS ^{13}CO data, a kinematic distance likelihood function may be computed for BGPS sources with a valid ^{13}CO spectrum. To maximize the use of kinematic information generated from this technique, we use the $T_{\rm source}(v)$ directly to compute $\mathcal{L}(v_{\rm LSR}, l, b; d_\odot)$, following the process described in EB13. Examination of Figure 3.3, however, reveals many small features in the extracted spectrum below the T_A threshold that are artifacts of the technique. To eliminate these extraneous bumps, which would appear as spurious small peaks in the DPDF, we mask $T_{\rm source}(v)$ to include only the region $v_{\rm LSR} \pm 3\sigma$ from the Gaussian fit to the primary peak before multiplying it by the rotation curve probability density function. For the example source in Figure 3.3, this preserves the small shoulder on the lower-velocity side of the main peak, appropriately widening the peaks in $\mathcal{L}(v_{\rm LSR}, l, b; d_\odot)$, while eliminating probability associated with the peak near $v_{\rm LSR} = 57$ km s^{-1}.

3.5 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 define a means for associating distances from literature catalogs with BGPS molecular cloud structures (§3.5.1) and demonstrate it with trigonometric parallax measurements of masers (§3.5.2) and robust KDA resolutions for H II regions (§3.5.3), both associated with sites of massive star formation.

3.5.1 Associating Molecular Cloud Structures with Literature Catalog Objects

3.5.1.1 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 $(\ell, b, v_{\mathrm{LSR}})$ 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 entires weighted by $1/\xi$, where the non-dimensional distance parameter is computed as

$$\xi = \left[\left(\frac{\Delta\theta}{\theta_a} \right)^2 + \left(\frac{\Delta v_{\rm LSR}}{\Delta v_a} \right)^4 \right]^{1/2} . \tag{3.2}$$

The quantities $\Delta\theta$ and $\Delta v_{\rm LSR}$ are the angular separation on the sky and velocity separation, respectively, between the continuum-identified source and the catalog object, and θ_a and Δv_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_a), so θ_a must be computed individually for each catalog object as $R_a/d_\odot$. 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 (3.2). Since association is not considered outside the volume defined by θ_a and Δv_a, the maximum possible value of ξ is $\sqrt{2}$. In the case of competing KDA resolutions for catalog objects with similar ξ, the resulting prior DPDF will be relatively unconstrained, accurately reflecting the distance uncertainty given the available information.

3.5.1.2 Optimization

As with the morphological spectrum extraction technique of §3.4, 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 bivariate cumulative distribution function of these quantities for the 749 GRS clouds studied by Roman-Duval et al. (2010) is shown as black contours in Figure 3.7, 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 90$^{\rm th}$ percentile values for

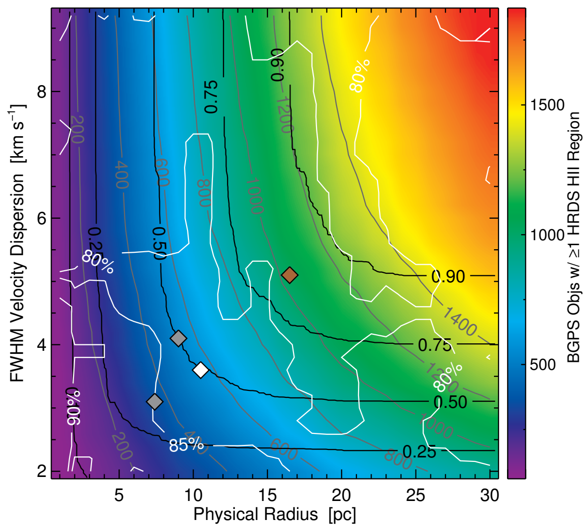

Figure 3.7 Optimization of the association volume for catalog-based priors, based on HRDS H II regions. Background color scale and gray contour denote the number of BGPS V2 objects falling within the association volume of one or more HRDS H II region, given the R_a and Δv_a indicated. Black contours show the bivariate cumulative distribution function for R and Δv 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 Δv_a while the brown and gray diamonds are discarded choices (see text).

each of R_a and Δv_a individually, but the combination lies along the 83$^{\rm rd}$ percentile contour.

Akin to the color scale of $N_{\rm comp}$ in Figure 3.5, the color scale and gray contours in Figure 3.7 depict the number of BGPS V2 sources associated with one or more HRDS H II region (§3.5.3). 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 DPDF$_{\rm hrds}$ (with association volume $R_a, \Delta v_a$) as a prior on $\mathcal{L}(v_{\rm LSR}, l, b; d_\odot)$ using Equation (3.1) with the distance derived using the priors DPDF$_{\rm emaf}$ and

$\mathrm{DPDF}_{\mathrm{H}_2}$ from EB13 in the same fashion. This comparison was done for 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-idenified 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 3.7. We used the $\mathrm{DPDF}_{\mathrm{emaf}}$ 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, as the random reassignment of KDA resolution between near and far (leaving tangent-point KDA resolutions unchanged) in the V2 distance catalog (see §3.6.4) yields a $\approx 64\%$ distance agreement.

As with optimizing the morphological spectrum extraction technique of §3.4, 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 50$^{\mathrm{th}}$ percentile contour of the bivariate cumulative distribution function, and settled on $(R_a, \Delta v_a) = (10.5\ \mathrm{pc}, 3.6\ \mathrm{km\ s}^{-1})$ as the best balance between distance agreement rate and number of objects included (white diamond). With these values, 17% 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-μm emission in $\mathrm{DPDF}_{\mathrm{emaf}}$. 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 $(0.92)^2 \approx 0.85$, not far from the rate indicated by the white diamond.

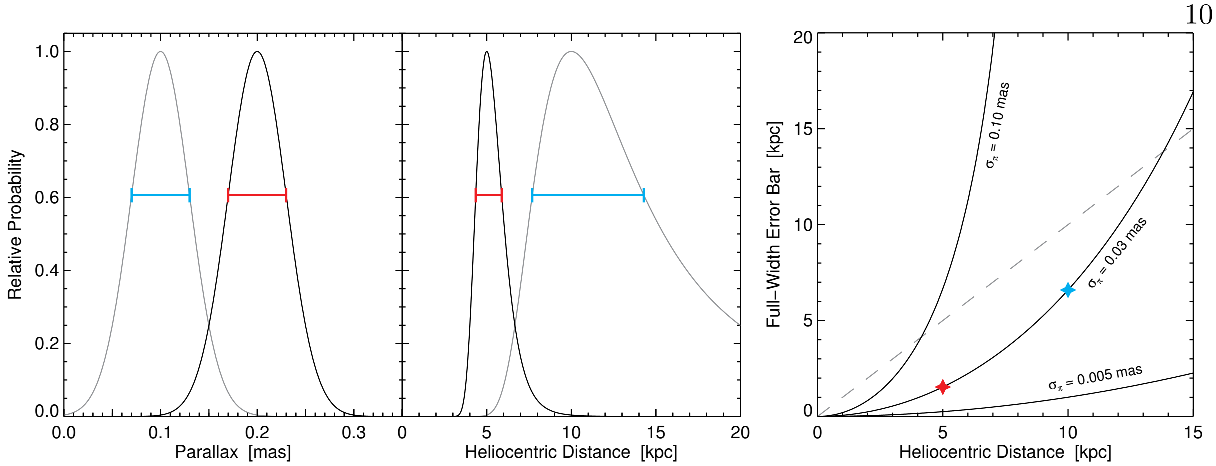

Figure 3.8 Example of $\mathrm{DPDF_{px}}$ construction from trigonometric parallax measurements. **Left**: Gaussians in parallax with $\mu_{\rm px} = 0.20$ mas ($\mu_{\rm px} = 0.10$ mas) and $\sigma_{\rm px} = 0.03$ mas are plotted in black (gray). The full-width error bars (**i.e.**, $\mu \pm \sigma$) 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 the 1:1 relationship.

Other diamonds in Figure 3.7 represent sub-optimal values. The brown diamond (mentioned above) has a relatively large association volume and encompasses over 1,000 BGPS objects, but is larger than 83% of GRS clouds and has a distance matching rate of less than 80%. The gray diamond at $(R_a, \Delta v_a) \approx (7.5\ \mathrm{pc}, 3\ \mathrm{km\ s^{-1}})$ lies at the intersection of the marginalized $50^{\rm th}$ percentile levels for each parameter (and so is reasonably matched to GRS cloud properties) and has an 85% distance matching rate, but is useful for less than 400 BGPS objects. Finally, the gray diamond along the chosen bivariate $50^{\rm th}$ percentile contour has a slightly higher matching rate than the white diamond, but encompasses fewer BGPS sources.

3.5.2 Trigonometric Parallax Measurements

The BeSSeL survey has been conducting VLBI observations of CH_3OH and H_2O masers associated with HMSFRs for the past several years (**cf.** 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 §3.3.2.2. For instance, the ≈ 16 km s^{-1} counterrotation of HMSFR G075.76+00.33 yields a kinematic distance of ≈ 5.7 kpc, apparently in the Perseus arm, but its parallax distance of 3.5 ± 0.3 kpc places it in the Local arm (Xu et al., 2013).

To construct the prior DPDF$_{\rm px}$ for trigonometric parallax measurements (π_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 $-12° \leq \ell \leq 241°$. Source ($\ell, b, v_{\rm LSR}$) 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 DPDF$_{\rm px}$ 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 DPDF$_{\rm px}$ for $\pi_s = 0.20 \pm 0.03$ (black) and $\pi_s = 0.10 \pm 0.03$ (gray) are shown in Figure 3.8. The typical parallax uncertainty in the present sample is ≈ 0.03 mas, which translates to a distance uncertainty of $^{+0.88}_{-0.65}$ kpc at $d_\odot = 5$ kpc, or a full-width error bar (see §3.6.3) of 1.53 kpc. The right panel of Figure 3.8 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 $d_\odot$.

3.5.3 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 H II Region Discovery Surveys (HRDS; Bania

Table 3.2. H II Region KDA Resolutions from the HRDS

Object Name	ℓ ($^\circ$)	b ($^\circ$)	$v_{\rm LSR}$ (km s^{-1})	$d_\odot$ (kpc)	KDA[a] Resol.	QF[b]	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

References. — 1: Anderson & Bania (2009); 2. Anderson et al. (2012), 3. Bania et al. (2012)

[a]N = near, F = far, T = tangent

[b]Quality Factor of the KDA resolution; see Anderson et al. (2012) for more details. Bania et al. (2012) does not assign a QF.

Note. — This table is available in its entirety in a machine-readable format in the online journal. A portion is shown here for guidance regarding its form and content.

et al., 2010, 2012) used the Green Bank Telescope 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 (**cf.** 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_2 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 $v_{\rm LSR}$ 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 $|v_{\rm HII}| < |v_{\rm absorption}| < |v_{\rm tan}|$ places the H II region at the far kinematic distance, and absorption only at $|v_{\rm absorption}| < |v_{\rm HII}|$ 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 3.2 as a reference for computing $\mathrm{DPDF}_{\rm hrds}$. Since the HRDS Galactic longitude range is $15° \leq \ell \leq 67°$, all objects with positive $v_{\rm LSR}$ must be assigned a KDA resolution (objects with $v_{\rm LSR} \leq 0$ km s^{-1} are unambiguously beyond the solar circle). Because this prior is based on the physically relevant tangent distance ($d_{\rm tan}$), we construct $\mathrm{DPDF}_{\rm hrds}$ as a step function removing probability on the opposite side of $d_{\rm tan}$. An analysis of posterior DPDFs suggests that objects within 1 kpc of $d_{\rm tan}$ should be given a "tangent" KDA resolution (§3.6.3), so a strict step function at $d_{\rm tan}$ would improperly bias the prior for these objects. We therefore model $\mathrm{DPDF}_{\rm hrds}$ with an error function possessing a rolloff width of 1 kpc for objects in Table 3.2 with a "N" or "F" resolution, and as a Gaussian centered on $d_{\rm tan}$ with $\sigma = 1$ kpc for objects with a "T" resolution.

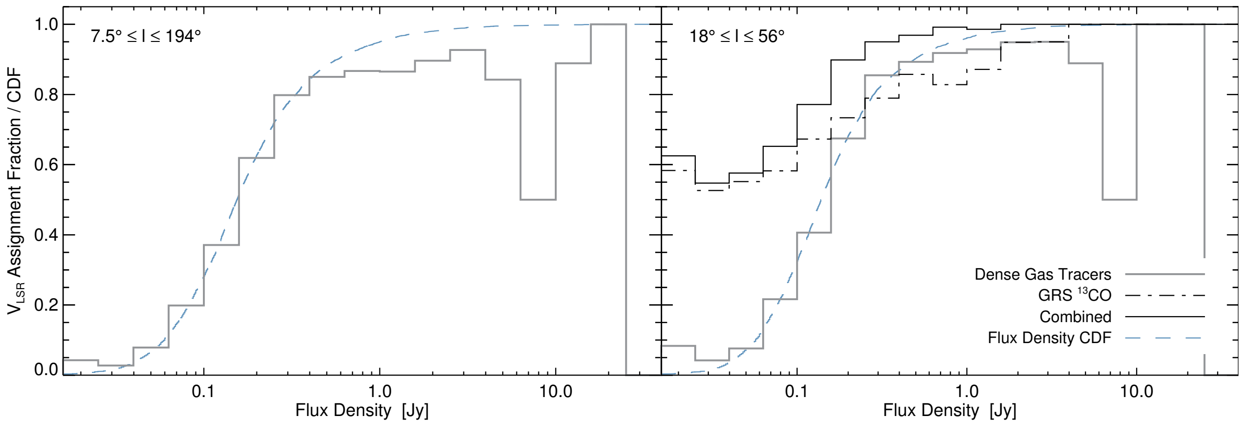

Figure 3.9 Assignment of $v_{\rm LSR}$ as a function of $\lambda = 1.1$ mm peak flux density. The blue dashed curve in each panel represents the cumulative distribution of flux densities for that subset of sources. **Left**: Velocity assignment fraction for the collection of dense-gas tracers discussed in §3.2.2 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}CO spectra (dot-dashed black) are shown, along with the combined velocity assignment fraction (solid black).

3.6 Results

3.6.1 Kinematics

3.6.1.1 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)[9] 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 $v_{\rm LSR}$ information with the V2.1 source catalog.

The Bolocat cataloging routine produces label maps that identify survey mosaic pixels be-

[9] The V2.1 catalog corrects various cataloging errors and represents the definitive catalog of objects in the V2.0 images (G13).

Table 3.3. Number of BGPS V2 Sources for Each Velocity Tracer

Species	N	Ref.
HCO^{+}(3-2)	2604	1
N_2H^{+}(3-2)	69	1
CS(2-1)	256	2
NH_3(1,1)	453	3,4
$C^{18}O$(2-1)	141	5
^{13}CO(1-0)	2660	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).

longing to each catalog entry (Rosolowsky et al., 2010b). To account for the finite solid angle encompassed by each spectroscopic pointing, any one $v_{\rm LSR}$ 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 $v_{\rm LSR}$ of spectra within the same survey (**i.e.**, HCO^+(3-2)) are compared. If the constituent velocities are more than 5 km s^{-1} discrepant, no $v_{\rm LSR}$ from that survey is assigned and a flag is returned, otherwise the spectral fit properties ($v_{\rm LSR}$, linewidth, uncertainties) are combined, weighted by the peak temperature of each spectrum. Second, the fit properties of different species are compared (**i.e.**, NH_3(1,1) vs. HCO^+(3-2)), and discrepant velocities result in no $v_{\rm LSR}$ 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 which have a valid $v_{\rm LSR}$ from each of the molecular line surveys used is presented in Table 3.3. As an example of the shift in kinematic information from version 1 of the BGPS catalog, the HHT surveys detected either HCO^+(3-2) or N_2H^+(3-2) emission associated with 3,126 BGPS version 1 sources, but only 2,676 version 2 sources have a valid (**i.e.**, not conflicting) $v_{\rm LSR}$ from these data. For the 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 3.3 add up to a larger number, a total of 2,925 BGPS V2 sources have a valid $v_{\rm LSR}$ from one or more of the dense gas spectral surveys discussed in §3.2.2.

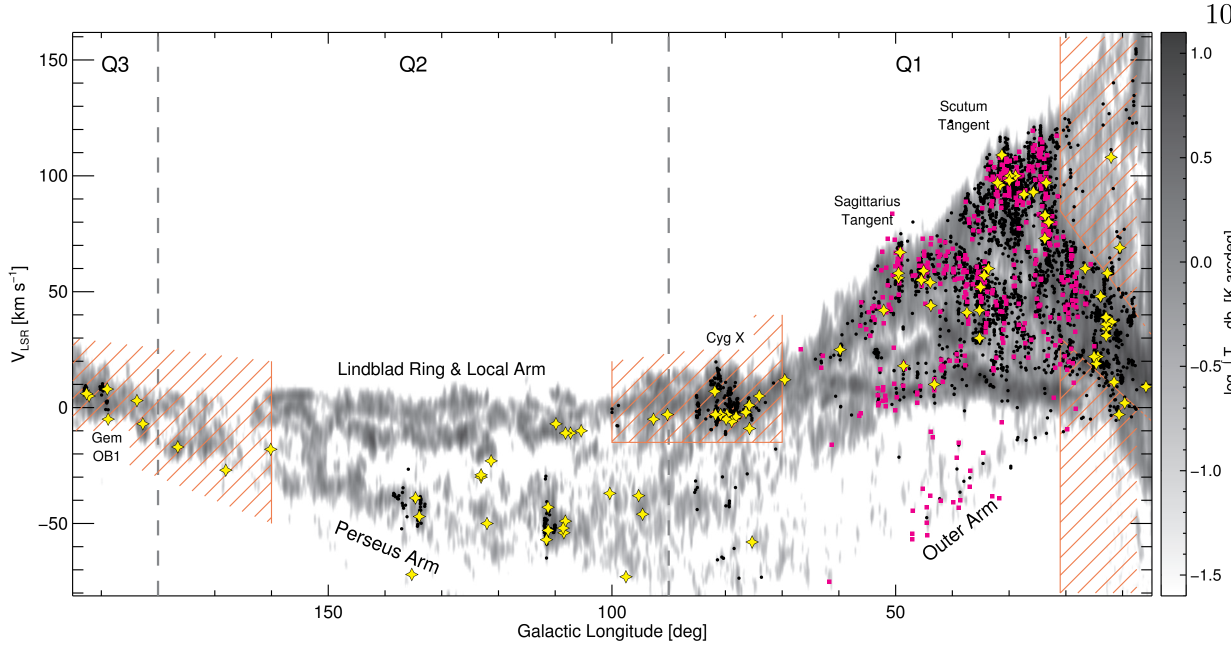

Figure 3.10 Longitude-velocity diagram for the BGPS. Background image is the latitude-integrated ^{12}CO intensity of Dame et al. (2001). Black circles mark the locations of BGPS V2 sources, and magenta squares identify the H II regions from Table 3.2. Yellow stars mark the trigonometric parallax measurements presented in Table 1 of Reid et al. (2014). The kinematic avoidance zones discussed in §3.3.2.2 are shown as hashed regions. Various Galactic features are identified.

3.6.1.2 Kinematic Catalog Expansion Using ^{13}CO

Because ^{13}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 Fig. 3.6). The molecular species described in §3.2.2 trace the dense environments of molecular cloud clumps and cores (Evans, 1999), so we preferentially use that information over the ^{13}CO data when it is available. New ^{13}CO kinematic distances are therefore limited to those objects without an assigned $v_{\rm LSR}$ from the dense-gas tracers due to non-detection, detection of multiple velocity components with $\Delta v \geq 5$ km s^{-1}, or sources not observed by spectroscopic surveys. As a result, while 2,660 Bolocat objects have valid ^{13}CO $v_{\rm LSR}$ (as defined by the flow chart of Figure 3.4) only 1,139 have a ^{13}CO velocity exclusively. The remainder overlap one or more dense gas tracer, and have velocities that agree with the other measurements 94% of the time. Combination of all available kinematic information yields a collection of 4,082 BGPS sources (representing 47% of the entire V2 catalog) whose velocity information is included with the expanded distance catalog (§3.6.4).

Shirley et al. (2013) showed a strong correlation between BGPS $\lambda = 1.1$ mm flux density and detection fraction in HCO^+(3-2) or N_2H^+(3-2). With the shift to BGPS V2 and the addition of ^{13}CO(1-0) as an additional velocity tracer, we reexamine this relationship. The fraction of sources in logarithmic flux density bins that have an assigned $v_{\rm LSR}$ are shown in Figure 3.9; sources without such an assignment may be spectroscopic non-detections, have multiple velocity components, or be unobserved. The left panel of Figure 3.9 reflects the Galactic longitude range of the Shirley et al. survey, the most comprehensive sample to date. The dips in the $v_{\rm LSR}$ assignment rate at $S_{1.1} \geq 4$ Jy are due to 11 catalog sources, 6 of which have multiple velocity components detected in HCO^+(3-2), two are new sources at the edge of BGPS mosaics not identified in the V1 catalog, and three 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 $v_{\rm LSR}$ assignment falls below 50% for $S_{1.1} < 200$ mJy still holds.

The right panel of Figure 3.9 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}CO spectra (dot-dashed black) is shown, along with the combined (**i.e.**, sources having a velocity from one or the other) assignment rate (solid black). Three features bear remark: (1) the velocity-assignment rate in ^{13}CO never falls below 50% even at low flux density, (2) the combined $v_{\rm LSR}$ assignment rate $\gtrsim 95\%$ for $S_{1.1} \geq 0.4$ Jy, and (3) all bright ($S_{1.1} \geq 4$ Jy) sources not assigned a dense-gas $v_{\rm LSR}$ have such an assignment from ^{13}CO. The kinematic distance likelihood provided by ^{13}CO is, therefore, a vital complement to the suite of dense-gas spectroscopic surveys.

3.6.1.3 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 3.10 illustrates the correlation between the more diffuse molecular component of the interstellar medium (latitude-integrated ^{12}CO of Dame et al., 2001, grayscale background) 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-Centarus 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 3$^{\rm rd}$ quadrant. The one feature in the ^{12}CO image not traced by dense gas is the plethora of molecular gas near $v_{\rm LSR} = 0$ km s^{-1} in the 2$^{\rm nd}$ 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 3.10 and generally follow the same patterns as the BGPS molecular cloud structures. The H II regions from Table 3.2 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 (§3.3.2.2).

3.6.2 Prior DPDFs

3.6.2.1 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 ($|b| \gtrsim 0\overset{\circ}{.}4$) to the near kinematic distance. This constraint derives from the relatively narrow thickness of the disk's molecular layer (half-width at half maximum ≈ 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 DPDF$_{\rm H_2}$ 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 $R_{\rm gal} \geq 0.2 R_0$. 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 $\mathrm{DPDF}_{\mathrm{H_2}}$ only skews the posterior DPDF to smaller $d_\odot$. We therefore limit the application of this prior to $11° \leq |\ell| \leq 70°$, with the upper limit corresponding to the start of the kinematic avoidance zone near $\ell = 90°$.

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; **cf.** Simon et al., 2006a), eight-micron absorption features (EMAFs) are visible as decrements in bright Galactic emission in the $\lambda = 8$ μm **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-μm emission (Robitaille et al., 2012) to construct a $\mathrm{DPDF}_{\mathrm{emaf}}$ 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 $\mathrm{DPDF}_{\mathrm{emaf}}$ was computed, 899 BGPS V2 sources met the final selection criteria, due in part to the expansion of the kinematic data set with GRS ^{13}CO data. Of this set of EMAFs, well-constrained distance estimates exist for 715 (see §3.6.4), an improvement in itself over the V1 set of 618 well-constrained sources from EB13.

3.6.2.2 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 301 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, 300 of these objects have well-constrained posterior DPDFs (see §3.6.3). 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 ~ 1.5 kpc discrepant from the parallax distance. Because $\mathrm{DPDF}_{\mathrm{px}}$ 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 232 have a $\mathcal{L}(v_{\mathrm{LSR}}, l, b; d_{\odot})$ despite all having a measured v_{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.

3.6.2.3 H II Regions

The combined HRDS catalog with robust KDA resolutions from Table 3.2 encompasses 441 H II regions, and a total of 564 BGPS V2 sources are associated with one or more of these regions. The plentiful nature of these objects (see Figure 3.10) means that, given the association volume defined in §3.5.1.2, a BGPS V2 source will occasionally ($\approx 14\%$) be associated with more than one HRDS object. Approximately 20% of this subset have an unconstrained $\mathrm{DPDF}_{\mathrm{hrds}}$ based on the conflict between KDA resolutions for constituent H II regions, accounting for some 3% of the 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 $\mathrm{DPDF}_{\mathrm{hrds}}$ and $\mathrm{DPDF}_{\mathrm{emaf}}$.

3.6.2.4 Gemini OB1

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 OB1 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 $\mathcal{L}(v_{\mathrm{LSR}}, l, b; d_{\odot})$. Fortunately, there exist VLBI trigonometric parallax measurements of H_2O and CH_3OH masers for three sources in this molecular cloud (shown as stars in Figure 3.11).

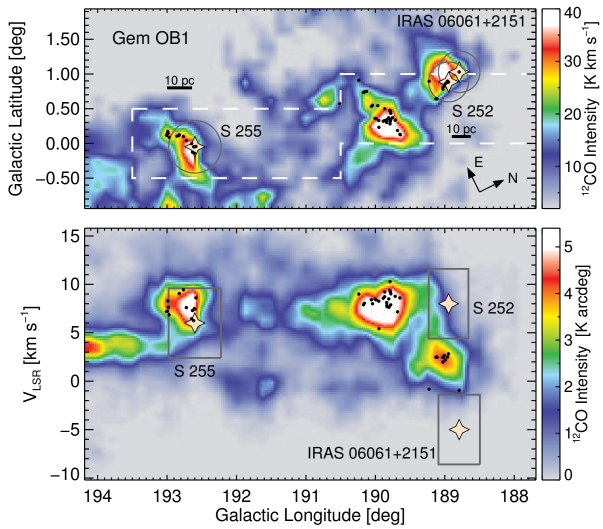

Figure 3.11 The Gemini OB1 field. **Top**: Sky-projected map of the Gemini OB1 molecular cloud. The background image is the velocity-integrated ^{12}CO data of Stacy & Thaddeus (1991); Dame et al. (2001), and black circles mark the locations of BGPS V2 sources. Stars mark the locations of the masers with trigonometric parallax measurements, and the enclosing circles mark the projected association volumes, computed as in §3.5.1. 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 $-1\overset{\circ}{.}0 \leq b \leq 1\overset{\circ}{.}5$ of the same field. Both panels share a common longitude axis.

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 3.11.

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}CO(1-0) emission in Figure 3.11, we assign the 1.59 kpc distance of S255 to the remainder of the BGPS sources in that 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}CO and dense-gas emission associated with the dust (Figure 3.11, **bottom**). The origins of these offsets are unclear, and it may be that these masers are associated with young

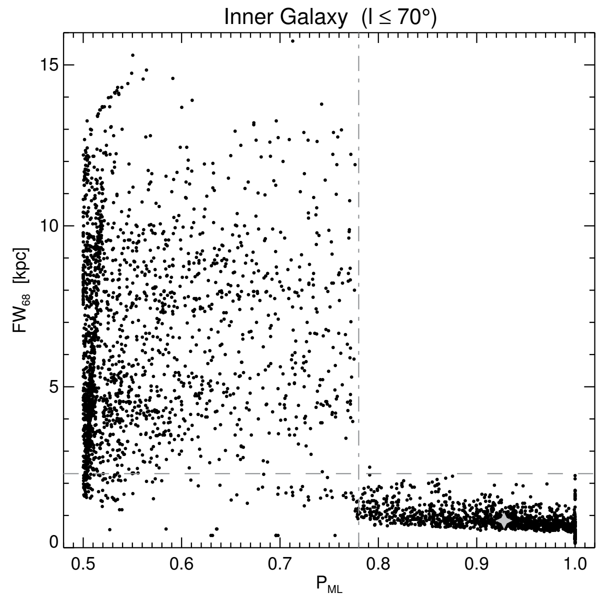

Figure 3.12 “Well-constrained” distance estimate criteria. Each BGPS V2.1 catalog object at $\ell \leq 70°$ is shown, plotted by its integrated posterior DPDF on the $d_{\rm ML}$ side of the tangent point ($P_{\rm ML}$) and the full-width error bar containing $\geq 68.3\%$ of the integrated posterior DPDF around $d_{\rm ML}$ (FW_{68}). The vertical dot-dashed line at $P_{\rm ML} = 0.78$ is shown to guide the eye, while the horizontal dashed line marks the $\mathrm{FW}_{68} \leq 2.3$ kpc boundary for a “well-constrained” distance estimate. The gray star identifies the median values of $P_{\rm ML}$ and FW_{68} for the well-constrained sample.

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 $\mathrm{DPDF}_{\rm px}$. 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 $\ell \lesssim 191°$.

3.6.3 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, $P_{\rm ML}$ is the integrated posterior DPDF on the side of the tangent point ($d_{\rm tan}$) containing the maximum-likelihood distance ($d_{\rm ML}$). For sources well away from $d_{\rm tan}$ 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, $0.5 \lesssim P_{\rm ML} \leq 1.0$. The measure of constraint tightness on a distance estimate is computed as the full-width of the region around $d_{\rm ML}$ containing at least 68.3% of the integrated probability and whose bounds occur at equal probability level (FW_{68}). Sources near $d_{\rm tan}$ will have $\mathrm{FW}_{68} \approx 1-2$ kpc due to localization from $dv_{\rm LSR}/d(d_\odot)$ (§3.3.2.2). For single-peaked DPDFs, 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.

Choice of criteria for "well-constrained" distance estimates once again relies on the bivariate distribution of $P_{\rm ML}$ and FW_{68}, shown in Figure 3.12. As in EB13, the pattern whereby FW_{68} becomes large for $P_{\rm ML} \lesssim 0.78$ holds. This break is due to the geometry of $\mathcal{L}(v_{\rm LSR}, l, b; d_\odot)$, 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 $P_{\rm ML} \gtrsim 0.78$, the maximum value of FW_{68} is approximately 2.3 kpc. We again choose to define a "well-constrained" distance estimate as $\mathrm{FW}_{68} \leq 2.3$ kpc to include objects within ≈ 1 kpc of $d_{\rm tan}$ (lower-left corner of Figure 3.12) because the near/far ambiguity for these sources leads to a distance discrepancy of $\lesssim 2$ kpc. We note that while we chose $\mathrm{FW}_{68} = 2.3$ kpc to define "well-constrained", it is clearly an upper limit, and the vast majority of these sources have $\mathrm{FW}_{68} \lesssim 1.5$ kpc, or a distance uncertainty of $\lesssim 0.8$ kpc. The median values of $P_{\rm ML}$ and FW_{68} for the well-constrained sample are 0.928 and 0.84 kpc, respectively (gray star in Figure 3.12).

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 $P_{\rm ML} = 1.0$ and FW_{68} well within the limit outlined above.

3.6.4 The V2 Distance Catalog and Posterior DPDFs

3.6.4.1 Catalog Description

A posterior DPDF for each BGPS V2.1 catalog object was computed using Equation (3.1)

Table 3.4. KDA Resolutions for BGPS V2 Sources

KDA Resolution	Flag	N_{kin} [a]	N_{tot} [b]	$f_{\mathrm{w.c.}}$ [c] (%)
Near	N	1036	1036	57.5
Far	F	353	353	19.6
Tangent	T	216	216	12.0
Outer Galaxy	O	197	197	10.9
Unconstrained	U	1887	6392	···
Excluded[d]	X	0	400	···
Total		3689	8594	100

[a]Number of objects in the "kinematic sample".

[b]Number of objects from the full Bolocat V2.

[c]Fraction of the well-constrained sources with this KDA resolution.

[d]Object lies in a kinematic avoidance zone (§3.3.2.2).

Table 3.5. BGPS V2 Distance Catalog

V2.1 Catalog Properties				Velocity		Heliocentric Distance				Galactocentric Position	
Catalog Number	ℓ (°)	b (°)	$S_{\rm int}$ (Jy)	$v_{\rm LSR}$ (km s^{-1})	Ref.	KDA Resol.	$P_{\rm ML}$ [a]	$d_{\rm ML}$ (kpc)	$\bar{d}$ (kpc)	$R_{\rm gal}$ [b] (kpc)	z [c] (pc)
3556	21.591	−0.147	1.02(0.15)	120.2	3	F	0.92	$9.50^{+0.26}_{-0.28}$	···	$3.53^{+0.14}_{-0.13}$	−26(5.2)
4386	26.091	−0.054	1.95(0.24)	28.0	1	F	0.94	$12.84^{+0.56}_{-0.50}$	···	$6.49^{+0.47}_{-0.41}$	−22(5.4)
4920	29.105	−0.108	0.26(0.07)	81.1	5	N	0.81	$4.52^{+0.44}_{-0.44}$	···	$4.91^{+0.26}_{-0.23}$	4.7(5.4)
5175	30.378	−0.005	0.43(0.09)	42.7	1,2	U	0.51	···	···	$6.17^{+0.26}_{-0.26}$	···
6046	34.481	0.126	0.17(0.05)	63.3	6	N	0.89	$3.70^{+0.42}_{-0.42}$	···	$5.69^{+0.24}_{-0.22}$	24(5.0)
6352	35.843	0.237	0.21(0.07)	29.2	6	U	0.55	···	···	$6.91^{+0.29}_{-0.29}$	···
6914	50.045	0.067	0.44(0.17)	54.5	6	T	0.50	$5.78^{+0.82}_{-1.66}$	5.36(1.14)	$6.39^{+0.10}_{-0.00}$	21(5.1)
6973	53.803	−0.162	0.19(0.11)	47.2	1,3	T	0.50	$4.92^{+0.82}_{-0.80}$	4.93(0.79)	$6.73^{+0.05}_{-0.00}$	2.3(6.2)
7884	111.646	0.388	0.67(0.07)	−54.6	1	O	1.00	$4.78^{+0.72}_{-0.68}$	···	$11.04^{+0.52}_{-0.47}$	63(7.4)
8218	192.824	0.125	0.54(0.15)	9.1	1,4	O	1.00	$1.60^{+0.06}_{-0.06}$	···	$9.91^{+0.06}_{-0.06}$	33(5.0)

[a] The integrated posterior DPDF on the $d_{\rm ML}$ side of the tangent point; used as a proxy for the quality of the distance constraint.

[b] $R_{\rm gal}$ is unambiguous and is computed from $v_{\rm LSR}$ and assumed uncertainty of 7 km s^{-1} (Reid et al., 2009b) for sources without a well-constrained distance estimate.

[c] Errors include contributions from variations in z along the line of sight over the range $d_\odot \pm \sigma_d$ and the ± 5 km s^{-1} 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).

Note. — Errors are given in parentheses.

Note. — This table is available in its entirety in a machine-readable format in the online journal. A portion is shown here for guidance regarding its form and content.

and the appropriate kinematic distance likelihood (dense gas or GRS ^{13}CO), and all applicable prior DPDFs. From the posterior DPDFs, the $P_{\rm ML}$ and FW$_{68}$ statistics, distance estimates and KDA resolutions were determined. The resulting distance catalog includes relevant information from the Bolocat V2.1, velocity information ($v_{\rm LSR}$, survey), and heliocentric distance and Galactocentric position, if available.

With the expansion of the source list beyond $\ell = 65°$, two new KDA resolution flags are introduced in this distance catalog. As in EB13, sources whose $d_{\rm ML}$ is within 1 kpc of $d_{\rm tan}$ 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 $d_{\rm ML}$ < (>) $d_{\rm tan}$ are again given the flag N (F). Outer-Galaxy objects, for which there is no KDA, are assigned O if they have an associated $v_{\rm LSR}$ 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 which either have no kinematic information or whose posterior DPDF does not meet the criteria of §3.6.3 are assigned the flag U. Table 3.4 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}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 3.5, which contains entries for each of the 3,689 catalog objects in the kinematic sample. Objects with a well-constrained distance estimate (flags N/F/T/O) have the maximum-likelihood distance ($d_{\rm ML}$) listed, along with the associated error bars. Tangent point objects additionally list the first-moment distance ($\overline{d}$), 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 $v_{\rm LSR}$, save those with KDA flag

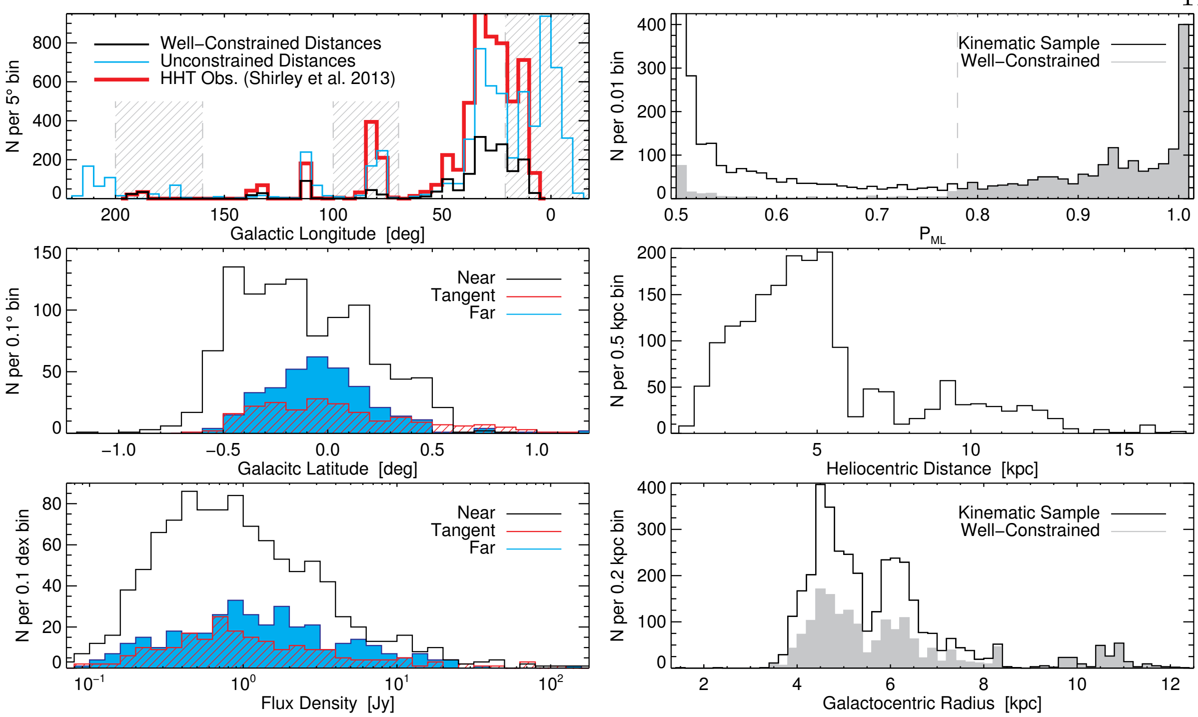

Figure 3.13 Summary of source properties from Table 3.5. **Top Left**: Comparison of the Galactic longitude distributions for objects with well-constrained (black) versus unconstrained (cyan) distances, with the red histogram showing the distribution of spectroscopic observations of Shirley et al. (2013). The gray hashed regions mark the longitude-projected kinematic avoidance zones (§3.3.2.2). **Middle Left**: Distributions of Galactic latitude for sources with well-constrained distance estimates at $\ell \leq 90°$. 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 $P_{\rm ML}$ 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).

X, as $R_{\rm gal}$ 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 8,594 objects in the BGPS V2.1 catalog are publicly available.[10]

3.6.4.2 Source Properties

[10] Available through IPAC at
`http://irsa.ipac.caltech.edu/data/BOLOCAM_GPS`

From the pool of sources in Table 3.5, 1,802 (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, 2014, to-be-submitted-in-the-very-near-future). A summary of the quantities in Table 3.5 is presented in Figure 3.13 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 (§3.3.2.2), although only the region around $\ell = 180°$ applies to all v_{LSR}. Any well-constrained objects within these zones either lie at an allowed v_{LSR} or are associated with a trigonometric parallax measurement (§3.5.2). 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 $\ell \geq 200°$, 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 $\ell \leq 90°$ in the middle- and bottom-left panels of Figure 3.13, 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 (FWHM $\approx 0\overset{\circ}{.}8$ and $\approx 0\overset{\circ}{.}9$, respectively) than the far kinematic group (FWHM $\approx 0\overset{\circ}{.}6$). This is to be expected as objects at $|b| \gtrsim 0\overset{\circ}{.}4$ are generally assigned the near kinematic distance by $\mathrm{DPDF_{H_2}}$. A 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 3.13 illustrate the distance resolution aspects of Table 3.5. The distribution of $P_{\rm ML}$ for the kinematic sample and subset of well-constrained sources is plotted in the top-right panel, with the overlap being complete for $P_{\rm ML} \geq 0.78$. The well-constrained sources with $P_{\rm ML} \approx 0.5$ are near the tangent point, as the peaks in the posterior DPDFs straddle $d_{\rm tan}$. The middle-right panel shows the heliocentric distance distribution of well-constrained sources, with 1284/1802 (71%) 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 $R_{\rm gal} \approx 5.5$ kpc, which may represent the division between the Scutum-Centarus and Sagittarius arms. Kinematic distances are unambiguous for $R_{\rm gal} > R_0$, so all objects in the kinematic sample beyond the solar circle have well-constrained distance estimates. The marked gap at $R_{\rm gal}$ =8.5 − 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 lies within a kinematic avoidance zone.

3.6.5 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 $(\ell, b, d_\odot) \rightarrow (R_{\rm gal}, \phi, z)$ conversion matrix from Appendix C of EB13, which accounts for the ≈ 25 pc vertical offset of the Sun above the Galactic midplane (Humphreys & Larsen, 1995; Jurić et al., 2008).

3.6.5.1 Face-On View of the Milky Way

The face-on map of the Milky Way from the north Galactic pole is shown in Figure 3.14,

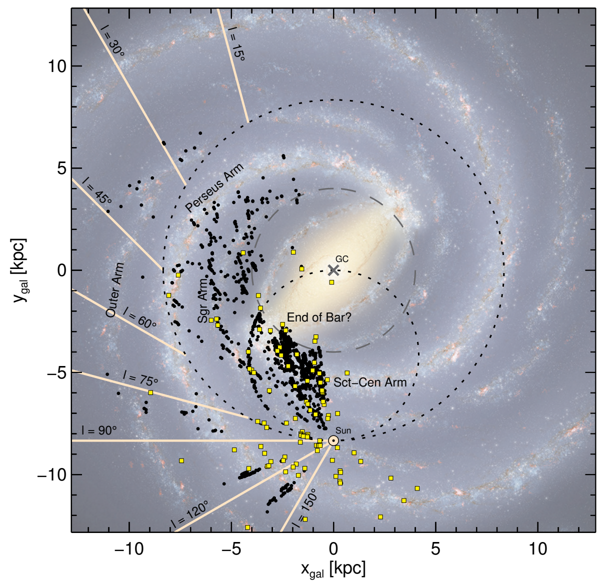

Figure 3.14 Face-on view of the Milky Way for sources with well-constrained distance estimates (black circles), plotted atop an artist's rendering of the Milky Way (R. Hurt: NASA/JPL-Caltech/SSC) viewed from the north Galactic pole. Yellow squares mark the locations of masers with trigonometric parallaxes (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 $R_{\rm gal} = 4$ 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. (A color version of this figure is available in the online journal.)

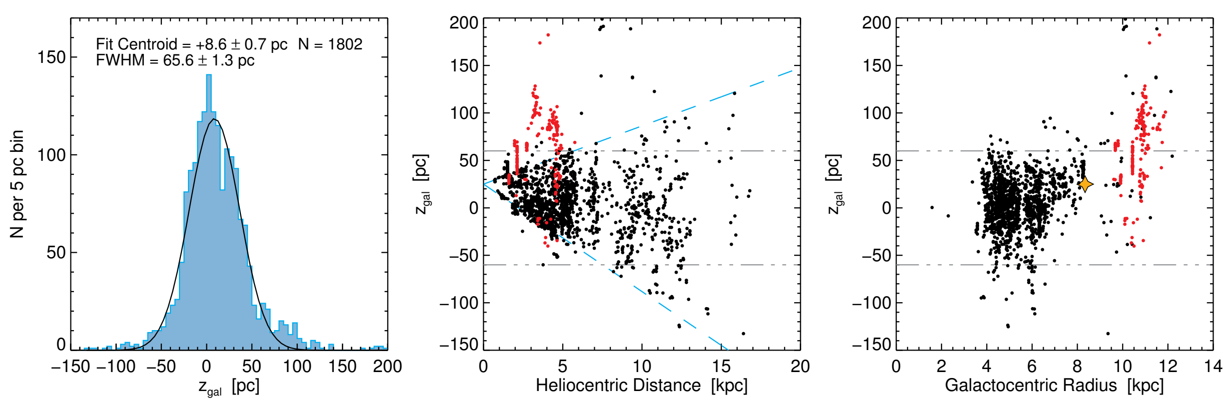

Figure 3.15 Vertical distribution of sources about the Galactic midplane. **Left**: Histogram of z with gaussian fit overplotted. **Center**: Vertical position as a function of heliocentric distance, with lines showing approximate boundaries of BGPS coverage at $\ell = 30°$. Sources plotted in red are at $\ell > 90°$. 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.

with the maximum-likelihood distance (or $\overline{d}$ for sources near $d_{\rm tan}$) 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 3.1. 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 $\ell = 30°$ and $\ell = 110°$). Error bars are not shown in Figure 3.14 for clarity, but with a typical uncertainty of ≈ 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 $v_{\rm LSR}$ 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.**, $\ell \approx 30° \pm 10°$, or $(x_{\rm gal}, y_{\rm gal}) \approx (-4, 0)$ 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 ≈ 80 BGPS sources in this area are primarily associated with HRDS H II regions; a (future, undeveloped) self-consistent H I absorption prior DPDF may solve this mystery.

Notwithstanding uncertainties in source location in Figure 3.14, 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 $\ell = 30°$ and the start of the Scutum-Centarus Arm moving to smaller longitude. Next is the general outline of the Sagittarius arm, visible from $(x_{\rm gal}, y_{\rm gal}) \approx$

$(-3, 3)$ kpc counterclockwise around to its tangency near $\ell = 50°$. Portions of the Perseus arm are traceable in the $\ell = 40° - 50°$ region and again in the outer Galaxy. Finally, the BGPS detects 23 objects in the Outer arm beyond the solar circle in the $\ell = 20° - 80°$ range, at a heliocentric distance of $\approx 10 - 15$ kpc.

3.6.5.2 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 3.15). The errors tabulated in the last column of Table 3.5 include contributions from variations in z along the line of sight over the range $d_\odot \pm \sigma_d$ and the ± 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 $+8.6 \pm 0.7$ pc, a FWHM of 65.6 ± 1.3 pc, and a reduced $\chi^2_{\mathrm{red}} = 1.8$. 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 $|b| \leq 0\overset{\circ}{.}5$ limit of BGPS coverage. The cyan dashed lines in that panel mark this limit at $\ell = 30°$ (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 ($\ell > 90°$) 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 (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. Use of the present DPDF methodology with ATLASGAL and/or Hi-GAL data (both of which extend to at least $|b| = 1°$) to derive distances to a large number of molecular cloud structures across the Galactic plane will yield a more complete picture of the vertical distribution of dense gas.

The rightmost panel in Figure 3.15 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. The gap in sources around $R_{\rm gal} = 5.5$ kpc is peculiar, and its source is not immediately clear. It may be a physical break between the Molecular Ring and more local star formation in the near Sagittarius arm, or an artifact of the chosen flat rotation curve.[11]

3.7 Discussion

3.7.1 ^{13}CO as 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 (**cf.** Dame et al., 2001). The low excitation density of the optically-thinner isotopologue ^{13}CO explains the many velocity components identified in surveys such as the GRS along any given line of sight, especially towards the crowded inner Galaxy. The catalog of molecular cloud clumps presented in Rathborne et al. (2009) implies that ^{13}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 $v_{\rm LSR}$ extracted directly from the ^{13}CO spectra of the GRS and that from a dense gas tracer (see §3.2.2), however, yields only a $\sim 85\%$ matching rate. This implies that while the ^{13}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

[11] Persic et al. (1996) derived a "universal" spiral galaxy rotation curve from the observed rotation curves of over 1,000 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 $R_{\rm gal} \lesssim 5$ kpc.

Table 3.6. Application of DPDFs to BGPS V2 Sources

DPDF	N	N_{wc}	Well-Constrained Sources Only Fraction of Method (%)	Fraction of w.c.[a](%)	Fraction of Full V2.1 (%)
Dense Gas v_{LSR}	2432	1262	51.9	70.0	14.7
^{13}CO v_{LSR}	1139	422	37.1	23.4	4.9
H_2 Scale Height	4474	1476	33.0	81.9	17.2
EMAF	899	715	79.5	39.7	8.3
Parallax Assoc.	301	300	99.7	16.6	3.5
HII Region Assoc.	564	536	95.0	29.7	6.2
Gemini OB1	49	49	100	2.7	0.6
Well Constrained	1802	1802	···	100	21.0

[a]Fraction of the set of well-constrained sources.

prior increases the agreement rate with HCO^+(3-2) to 94%, with approximately one in eight of the disagreeing sources having a v_{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_{eff} into a powerful dense gas tracer.

The morphological spectrum extraction technique is able to assign a v_{LSR} to molecular cloud structures at a high rate, exceeding 80% for $S_{1.1\ \mathrm{mm}} \gtrsim 0.4$ Jy, and does significantly better than dense gas tracers for $S_{1.1\ \mathrm{mm}} \lesssim 0.1$ 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 Galactic Ring Survey 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}CO, ^{13}CO, and C^{18}O over the range $305° \leq \ell \leq 345°$ and $|b| \leq 0.5°$ with ~ 0.7 K rms sensitivity. With $35''$ angular and 0.1 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}CO(3-2) survey (Dempsey et al., 2013) will cover $10° \leq \ell \leq 65°$ and $|b| \leq 0.5°$

Table 3.7. Well-Constrained Fraction for Overlapping Priors

	H_2 Scale Height	EMAF	Parallax Assoc.	HII Region Assoc.	Gemini OB1
H_2 Scale Height	0.330	0.790	0.995	0.950	···
EMAF		0.795	1.000	0.855	···
Parallax Assoc.			0.997	1.000	···
HII Region Assoc.				0.950	···
Gemini OB1					1.000

with ~ 1 K rms sensitivity and (smoothed) $16''$ angular and 1 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 3.5.

3.7.2 Inter-Comparison of Prior DPDF Methods

The performance of each DPDF method presented here is shown in Table 3.6. 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 the remaining columns use the number of well-constrained sources ($N_{\rm wc}$) as the numerator to indicate the constituent fractions of various larger sets. The "fraction of method" column ($N_{\rm wc}/N$) 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 3.6 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 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 3.7 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 3.6. 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_2 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 4,474 objects with $\mathrm{DPDF}_{\mathrm{H}_2}$, some 3,275 (73%) are in the kinematic sample, placing an upper limit on the KDA resolution rate. Of this kinematic pool, only 1,476 (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 (**cf.** Battersby et al., 2010; Dunham et al., 2011a), whereas EMAFs correspond to the earlier stages (EB13). Ellsworth-Bowers et al. noted that bright PAH emission near $\lambda = 8\ \mu$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 $\mathrm{DPDF}_{\mathrm{emaf}}$. 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-μm 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 $\mathrm{DPDF}_{\mathrm{px}}$ should be expected, as the precision of recent VLBI trigonometric parallaxes produces very narrow distance uncertainties. The sole $\mathrm{DPDF}_{\mathrm{px}}$ out-

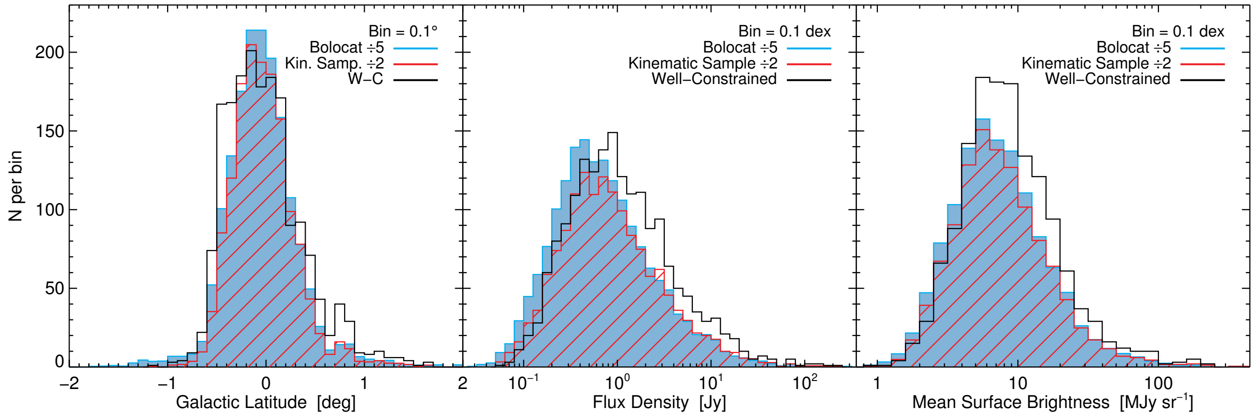

Figure 3.16 Comparison of observable quantities for the entire BGPS catalog (divided by 5, blue), the kinematic sample (divided by 2, red), and set of well-constrained distance estimates (black). **Left**: The Galactic latitude distributions. **Middle**: Source $\lambda = 1.1$ mm flux density distributions. **Right**: Source-averaged surface brightness, computed using Equation (3.3).

lier 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 §3.5.1 does not exceed the physical association scale length of HMSFRs in the Galactic plane.

3.7.3 A Representative Sample?

The 1,802 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 Bolocat, the kinematic sample of 3,689 sources, and the set of well-constrained distance estimates are shown in Figure 3.16 for the distributions of Galactic latitude, flux density, and mean surface brightness. The entire Bolocat (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

$$SB = \frac{S_{1.1}}{\pi\theta_R^2} = 12.4 \text{ MJy sr}^{-1} \left(\frac{S_{1.1}}{1 \text{ Jy}}\right) \left(\frac{\theta_R}{33''}\right)^{-2} \quad , \tag{3.3}$$

where $S_{1.1}$ is the BGPS integrated flux density and θ_R is the deconvolved radius of the catalog source (Rosolowsky et al., 2010b). 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 $b \approx -0.1°$, and a FWHM $\approx 0.7°$, 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, 0.74 Jy, and 0.93 Jy, respectively. This skew for the kinematic sample is accounted for in the kinematic detection rate as a function of flux density (Figure 3.9). 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, 7.8, and 8.0 MJy sr^{-1} 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 99.997% confidence level or greater, but the full catalog and kinematic samples cannot be distinguished at the 90% 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 θ_R is $0.75 - 0.79$ for all three groups; see Equation 3.3) 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 towards brighter and denser objects than either the Bolocat as a whole or even the kinematic sample of objects with detected v_{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_3(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.

3.8 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^+(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 ^{13}CO data of the GRS for molecular cloud structures using the dust-continuum data as prior information. The low effective density for excitation of ^{13}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_{LSR}. When compared to using the v_{LSR} of the brightest ^{13}CO peak along the line of sight, this morphological spectrum extraction technique increases the velocity-matching rate between ^{13}CO v_{LSR} and HCO^+(3-2) from $\approx 85\%$ to $\approx 94\%$. The additional velocities derived from ^{13}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 301 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 $\ell = 180°$ 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 3,689 BGPS V2 sources with kinematic information, 1,802 (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-Centarus arm between the Sun and the Galactic center. The vertical distribution of BGPS molecular cloud structures is narrower than the measured distribution of CO in the disk, but our measurement should be viewed as a lower limit due to the restricted latitude coverage of the BGPS. Finally, while the collection of well-constrained distances represents only 21% of the BGPS catalog, the distributions of observable quantities are generally representative of the entire catalog, but are biased towards higher-density sources due to the suite of prior DPDFs currently available.

Chapter 4

Physical Properties and Mass Functions of Dense Molecular Cloud Structures

We use the distance probability density function (DPDF) formalism of Ellsworth-Bowers et al. to derive physical properties for the collection of 1,802 Bolocam Galactic Plane Survey (BGPS), version 2, sources with well-constrained distance estimates. To account for Malmquist bias, we estimate that the present sample of BGPS sources is 90% complete above 400 $M_\odot$ and 50% complete above 75 $M_\odot$. The mass distributions for the entire sample and astrophysically motivated subsets are generally well-fit by a lognormal function, with approximately power-law distributions at high mass, and the power-laws emerge more clearly when the sample population is narrowed in heliocentric distance. The high-mass end power-law indices are generally $1.85 \leq \alpha \leq 2.05$, intermediate between that of giant molecular clouds and the stellar initial mass function. The fit to the entire sample yields a high-mass power-law $\hat{\alpha} = 1.94^{+0.16}_{-0.12}$. The physical properties of BGPS sources are consistent with large molecular cloud clumps or small molecular clouds, but the fractal nature of the dense interstellar medium makes difficult the mapping of observational categories to the dominant physical processes driving the observed structure. We present a fit to the mass-size Larson scaling relationship that extends the $M \propto R^{\approx 2}$ fit a further order of magnitude in mass beyond previous measurements. The face-on map of the Galactic disk's mass surface density based on BGPS dense molecular cloud structures reveals the massive star-forming regions W43, W49, and W51 as prominent mass concentrations in the first quadrant. The mass-weighted vertical distribution of dense gas has a scale height ($\approx$ 30 pc) consistent with the BGPS number-count distribution, and is approximately half that of the ^{12}CO layer. We present an 0.25-kpc resolution

map of the dense gas mass fraction across the Galactic disk that peaks around 5%.

4.1 Introduction

As the possible precursors to stellar clusters, OB associations, or smaller stellar groups, molecular cloud clumps and cores have become a primary focus for understanding the process of massive star formation (**cf.** McKee & Ostriker, 2007, and references therein). The recent advent of large-scale 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., 2010b) have detected tens of thousands of these objects in thermal dust emission. The true payoff of blind surveys in this portion of the electromagnetic spectrum is deriving the physical properties of regions hosting massive star formation. A detailed census of these dense molecular cloud structures can help constrain star-formation and galactic-evolution theories (**cf.** Kennicutt & Evans, 2012). Studies of their physical properties (**cf.** Peretto & Fuller, 2010; Giannetti et al., 2013) and mass distributions (**cf.** Netterfield et al., 2009; Olmi et al., 2013; Gómez et al., 2014) of these objects have begun, but despite the richness of the current data sets, a coherent picture has not yet emerged for the evolution of the dense interstellar medium and the origin of the stellar initial mass function.

While giant molecular clouds (GMCs) have been studied for several decades using the lowest rotational transitions of the isotopologues of CO and other simple molecules (**e.g.**, Scoville & Solomon, 1975; Cohen et al., 1980; Dame et al., 1987, 2001, and references therein), and studies of the apparently uniform stellar and cluster initial mass functions (**e.g.**, Bastian et al., 2010) go back to Salpeter (1955), the observational technology (detectors and angular resolution) has only recently developed to study the dense substructures intermediate between these two extremes. Furthermore, studies of nearby molecular cloud complexes have recently yielded estimates of the core mass function (**e.g.**, Alves et al., 2007; Swift & Beaumont, 2010) for the progenitors of single stellar systems, but it has only been with the recent large-scale blind surveys that the study of the **clump** mass function has been possible.

Theoretical modeling of molecular cloud structure evolution is beginning to place constraints on the clump mass function (**cf.** Donkov et al., 2012; Veltchev et al., 2013), and two primary functional forms are discussed in the literature: the power-law and lognormal distributions. Both arise from physical processes in molecular clouds; supersonic turbulence within molecular clouds produces a lognormal density distribution (**cf.** Padoan et al., 1997, although Tassis et al., 2010 find other means to produce a lognormal), while gravitational collapse of dense structures tends to produce a power law (**cf.** Padoan & Nordlund, 2002); the distinctions between these have implications for competing theories of massive star formation (Elmegreen, 1985). Furthermore, the observed mass function for dense molecular cloud structures should resemble a combination of these forms due to the complex interaction of physical processes (**cf.** Offner et al., 2013), and for a discussion of the details, see Hopkins (2013).

In a broader context, the Galactic distribution of dense molecular gas and star formation has implications for using the Milky Way as ground truth for studies of extragalactic star formation and galaxy formation. Cosmological simulations of galaxy formation and evolution have made remarkable progress in the last two decades reproducing some of the observed properties of galaxies (albeit with some notable exceptions, such as reproducing the stellar-to-dark-matter mass ratios across the entire mass spectrum of galaxies). Indeed, by incorporating feedback from stars, numerical galaxy evolutions models can now reliably produce spiral galaxies. However, "subgrid" physics recipes are used for physical processes that cannot be resolved in the simulations (**e.g.**, Kim et al., 2014). Derived physical properties of dense molecular cloud structures in the Milky Way can place constraints on masses and pressures, possibly providing a foundation for scaling relationships for star formation rates with respect to more active galaxies (Kamenetzky et al., 2014).

As the first completed dust-continuum millimeter-wavelength survey of the northern Galactic plane, the Bolocam Galactic Plane Survey (BGPS) has provided a foundation for studying dense molecular cloud structures in a variety of Galactic environments. The majority of objects detected in the BGPS correspond to molecular cloud clumps (Dunham et al., 2011b), which will likely form a cluster of stars. Derivation of physical properties of these objects requires robust distance estimates; Ellsworth-Bowers et al. (2013); Ellsworth-Bowers et al. (2014) developed a distance probability density function (DPDF) methodology for utilizing all available information towards this goal. By combining kinematic distance information with prior DPDFs based on ancillary data (such as **Spitzer**/GLIMPSE mid-infrared images and trigonometric parallax measurements) in a Bayesian framework, Ellsworth-Bowers et al. (2014, hereafter EB14) presented a catalog of 1,802 BGPS objects with well-constrained distance estimates.

In this work, we build on that distance catalog to derive the physical properties of and fit mass functions to this vital population of sources. The probabilistic description of the distance to each object in the survey catalog may be used to directly propagate uncertainty into the derived physical properties via Monte Carlo methods. We present here a catalog of physical properties for this set of sources, and fit mass functions to the entire sample, as well as astrophysically motivated subsets. Furthermore, we explore the extension of fundamental molecular cloud scaling relationships (Larson, 1981) to dense substructures and map out the distribution of dense molecular gas throughout the Galactic disk.

This paper is organized as follows. Section 4.2 describes the data used. The computation of physical properties from the DPDFs and construction of mass functions are described in Section 4.3. Section 4.4 presents the results, and implications of this work are discussed in Section 4.5. A summary is presented in Section 4.6.

4.2 Data

The Bolocam Galactic Plane Survey version 2 (BGPS V2; Aguirre et al., 2011; Ginsburg et al., 2013, hereafter G13), is a $\lambda = 1.1$ mm continuum survey covering 192 deg^2 at $33''$ resolution. It is one of the first large-scale blind surveys of the Galactic plane in this region of the spectrum, covering $-10° \leq \ell \leq 90°$ with $|b| \leq 0\overset{\circ}{.}5$ and selected larger cross-cuts to $|b| \leq 1\overset{\circ}{.}5$, plus selected regions in the outer Galaxy. For a map of BGPS V2 coverage and details about observation methods and the data reduction pipeline, see G13. From the BGPS V2 images, 8,594 millimeter dust-continuum sources were identified using a custom extraction pipeline. BGPS V2 pipeline products, including image mosaics and the catalog, are publicly available.[1]

The BGPS data pipeline removes atmospheric signal using a principle component analysis technique that discards common-mode time-stream 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 sky-subtraction behaves like a high-pass filter that 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).

Ellsworth-Bowers et al. (2014) presented a distance catalog for BGPS V2 sources using a Bayesian DPDF framework. The DPDFs describe the relative probability of finding an object at points along the line of sight and are constructed from multiple sources of distance-related information via

$$\mathrm{DPDF}(d_\odot) = \mathcal{L}(v_{\mathrm{LSR}}, l, b; d_\odot) \prod_i P_i(l, b; d_\odot) \ , \tag{4.1}$$

where $\mathcal{L}(v_{\mathrm{LSR}}, l, b; d_\odot)$ is the kinematic distance likelihood function based on a line-of-sight velocity ($v_{_\mathrm{LSR}}$), and the $P_i(l, b; d_\odot)$ are prior DPDFs based on ancillary data. Given circular orbits about the Galactic center, there are two points along a line of sight looking interior to the Solar circle that correspond to a single $v_{_\mathrm{LSR}}$, the kinematic distance ambiguity (KDA). The $\mathcal{L}(v_{\mathrm{LSR}}, l, b; d_\odot)$ for such objects is double-peaked, representing the equal probability of finding the object at either kinematic distance; prior DPDFs are used principally to resolve the KDA, placing an object at either the near or far distance (see EB14 for further discussion). The resulting catalog contains 1,802 sources with well-constrained distance estimates (distance uncertainties generally $\lesssim 0.5$ kpc), hereafter referred to as the Distance Catalog. The posterior DPDFs[2] are normalized to unit integral probability ($\int_0^\infty \mathrm{DPDF}\ \mathrm{d}d_\odot = 1$) to facilitate marginalization over distance, and allow for the Monte-Carlo propagation of uncertainty in an object's distance to the derived physical properties by randomly

[1] Available through IPAC at
`http://irsa.ipac.caltech.edu/data/BOLOCAM_GPS`

[2] Available with the BGPS release.

sampling the DPDFs many times.

4.3 Method

4.3.1 Computing Physical Properties Using the DPDF

It is straightforward to compute physical properties of dense molecular cloud structures using single-value distance estimators. The complete information contained in an object's posterior DPDF, however, represents the combination of knowledge and ignorance about its distance based on a wide collection of data encapsulated in the suite of prior DPDFs applied. Use of this full information, therefore, requires careful consideration. Below, we discuss mass calculation for BGPS objects in some detail (§4.3.1.1), and then apply those methods to the computation of other physical properties (§4.3.1.2).

4.3.1.1 Mass Derivation

A simple estimate of a dense molecular cloud structure's mass may be computed from optically thin millimeter dust continuum data via

$$M = \frac{\mathsf{r}\ S_{\nu,\mathrm{int}}\ d_\odot^2}{\kappa_\nu\ B_\nu(T_d)}\ , \tag{4.2}$$

where $S_{\nu,\mathrm{int}}$ is the source-integrated flux density, $d_\odot$ is the estimated heliocentric distance, κ_ν is the opacity of dust, $\mathsf{r} \equiv (m_{\mathrm{g}}/m_{\mathrm{d}})$ is the gas-to-dust mass ratio, and $B_\nu(T_d)$ is the Planck function evaluated at dust temperature T_d. For the specific case of the BGPS,

$$M = 13.1\, M_\odot\ \left(\frac{e^{13.0\ \mathrm{K}/T_d} - 1}{e^{13.0\ \mathrm{K}/20.0\ \mathrm{K}} - 1}\right)\left(\frac{S_{1.1}}{\mathrm{Jy}}\right)\left(\frac{d_\odot}{\mathrm{kpc}}\right)^2\ , \tag{4.3}$$

where $S_{1.1}$ is the $\lambda = 1.1$ mm source-integrated flux density, $\kappa_\nu = 1.14$ cm^2 g^{-1} of dust (Ossenkopf & Henning, 1994, Table 1, Column 5),[3] $\mathsf{r} = 100$ (Hildebrand, 1983), and the equation has been normalized to $T_d = 20$ K. Standard error-propagation methods may be used to determine the mass

[3] This table of κ_ν represents dust grains with ice mantles, coagulating in cold, dense molecular regions at $n = 10^5$ cm^{-3} for 10^6 yr. This is appropriate for molecular cloud **cores**, and may be an upper limit for less-dense **clumps** (**cf.** Martin et al., 2012).

uncertainty, but the asymmetric nature of many DPDF error bars (EB14) complicates the issue. The uncertainty in the distance may be marginalized over to produce the expectation value of the mass via

$$\langle M \rangle = \frac{\mathsf{r}\ S_{\nu,\mathrm{int}}}{\kappa_\nu\ B_\nu(T_d)} \int_0^\infty \mathrm{DPDF}\ d_\odot^2\ \mathrm{d}(d_\odot)\ , \tag{4.4}$$

where $\langle M \rangle$ is proportional to the second moment of the DPDF. The bimodal nature of many posterior DPDFs in the inner Galaxy, however, causes this to be a biased estimator of the mass, in much the same way $\overline{d}$ (the first-moment distance) can be a biased estimator of the distance. For instance, an object with a well-constrained distance estimate may have up to $\approx 20\%$ of the integrated DPDF in the non-favored kinematic distance peak; the difference between $\overline{d}$ and d_{ML} can be greater than 1 kpc, or $\approx 2\times$ the uncertainty in d_{ML} (see Ellsworth-Bowers et al., 2013 for a complete discussion).

This bias, coupled with the measured source flux density and assumed dust temperature uncertainties suggest using a Monte Carlo approach for estimating an object's mass and associated uncertainty. To determine the maximum-likelihood mass for a given catalog object, Equation (4.2) is computed many times with each realization drawing a distance, flux density, and (optionally) a temperature from the suitable probability density functions. The resulting (properly normalized) mass probability density function (MPDF) may then be used to estimate M_{ML} and the 68.3% error bars in the same fashion as done with DPDFs in EB14.

The MPDF, being the probability of an object having a mass between M and $M + \delta M$, is created from the histogram of masses produced by many ($\sim 10^6$) Monte Carlo realizations. As with any process involving binned data, choice of binning scheme can have profound effects on the derived quantities (Rosolowsky, 2005). Note that the computation of DPDFs is done analytically with respect to a fixed linear distance scale and is not subject to the need to create histograms from Monte Carlo calculations. In light of the spread of the possible masses for a single object and the dynamic range of integrated flux densities for BGPS sources measured in orders of magnitude, we chose to use logarithmic binning for computation of MPDFs and the probability density functions for other physical properties.

4.3.1.2 Physical Radius and Number Density

In addition to the masses of dense molecular cloud structures, the catalog of heliocentric distances enables the computation of physical sizes and densities of these objects. Not only can these physical properties help to disentangle the various populations of objects (**i.e.**, cloud, clump, core) detected by the BGPS and others, but they can also test the extension of various empirical relationships for GMCs down to more dense substructures.

The physical radius is computed from

$$R = 0.29 \text{ pc } \left(\frac{\theta_R}{\text{arcmin}}\right)\left(\frac{d_\odot}{\text{kpc}}\right) , \tag{4.5}$$

where θ_R is the deconvolved angular radius of a detected object computed from geometric mean of the deconvolved major (σ_{maj}) and minor (σ_{min}) axes of the flux density distribution,

$$\theta_R = \eta \left[\left(\sigma^2_{\text{maj}} - \sigma^2_{\text{bm}}\right)\left(\sigma^2_{\text{min}} - \sigma^2_{\text{bm}}\right)\right]^{1/4} \tag{4.6}$$

(Rosolowsky et al., 2010b). For the BGPS, the rms size of the beam is given by $\sigma_{\text{bm}} = \theta_{\text{FWHM}}/\sqrt{8 \ln 2} = 14''$, and $\eta = 2.4$ is a factor relating the rms size of the emission distribution to the angular radius of the object.[4] Objects whose measured minor axis dispersion is smaller than σ_{bm} do not have a finite θ_R; real objects should not fall into this category, but effects of the cataloging process lead to 29% of the full BGPS V2.1 catalog and 21% of the Distance Catalog having non-finite θ_R. These objects are excluded from analysis of physical radius and number density in §4.4.

As a means for classifying dense molecular cloud structures, the mean number density offers possible insight into the physical processes, such as structure formation via turbulence or gravitational collapse, that may be at play in a given catalog object. The number density com-

[4] The appropriate value of η to be used depends on the true emission distribution of the object and its size relative to the beam. Rosolowsky et al. (2010b) computed η for a large range of models, varying the emissivity distribution, angular size relative to the beam, and signal-to-noise ratio. The chosen value for the BGPS catalog is the median value from these simulations, but the variations span more than a factor of two. Use of a specific source model in conjunction with the BGPS catalog would require the appropriate value of η and rescaled catalog values.

puted with assumed cylindrical geometry (**i.e.**, a circle on the sky with radius R, and depth R),[5] $n = M/\pi\mu m_H R^3$, may be parameterized in terms of mass and radius or observable quantities as

$$\begin{aligned} n &= 4.60 \times 10^2\ \mathrm{cm}^{-3} \left(\frac{M}{100\,M_\odot}\right)\left(\frac{R}{\mathrm{pc}}\right)^{-3} \\ &= 2.47 \times 10^3\ \mathrm{cm}^{-3} \left(\frac{e^{13.0\ \mathrm{K}/T_d} - 1}{e^{13.0\ \mathrm{K}/20.0\ \mathrm{K}} - 1}\right)\left(\frac{S_{1.1}}{\mathrm{Jy}}\right)\left(\frac{\theta_R}{\mathrm{arcmin}}\right)^{-3}\left(\frac{d_\odot}{\mathrm{kpc}}\right)^{-1}. \end{aligned} \tag{4.7}$$

The first parameterization relates the physical quantities in units appropriate to dense molecular cloud structures, while the second is specific to the observational quantities of the BGPS and the assumptions used in Equation (4.3). Note that the number density is not as strong a function of heliocentric distance as is the mass; incorrect KDA resolution will have a smaller effect on the derived value.

4.3.2 Constructing Mass Distributions and Estimating Completeness

The mass function describing the relative numbers of dense molecular cloud structures is of particular interest for connecting star-formation theory to observation. Functional forms are fit to the distribution of masses and we explore some of the possible mass distribution constructions. Additionally, since the range of masses over which conclusions may be drawn is restricted by completeness considerations, we present an analysis estimating the mass completeness levels for a heterogenous data set such as the BGPS.

4.3.2.1 Mass Distributions

In its simplest form, the mass distribution of dense molecular cloud structures may be compiled from the direct application of Equation (4.2) to each object with a well-constrained distance estimate, where $d_\odot$ is either the maximum-likelihood distance or the first-moment distance, as discussed in EB13. This form of the distribution for the Distance Catalog sources is shown as the

[5] This parameterization of the volume results in $V = \pi R^3$; assuming spherical geometry increases the volume by a factor of 4/3, and assuming a cylinder of depth $2R$ increases it by a factor of 2. As shown in §4.4.1, the number density is a poorly constrained quantity for these objects, and source-to-source geometry differences play a significant role.

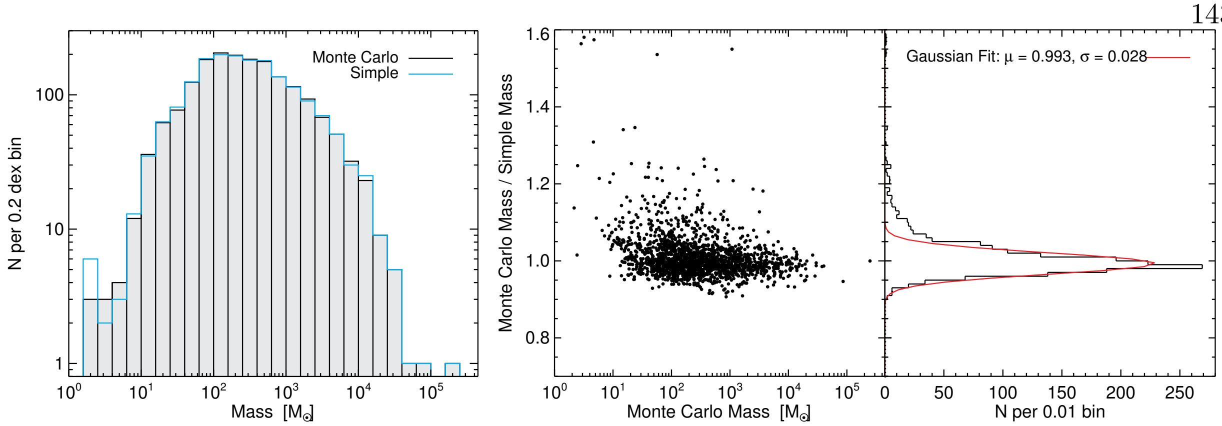

Figure 4.1 Comparison of mass distribution methods. **Left**: The mass distribution computed by applying Equation (4.2) to each object, using the maximum-likelihood distance (or the first-moment distance, as discussed in EB13) from the posterior DPDF ('Simple', cyan) plotted atop the mass distribution using the maximum-likelihood masses from the MPDFs ('Monte Carlo', black). **Middle**: Ratio of the 'Monte Carlo' mass to the 'Simple' mass as a function of 'Monte Carlo' mass. **Right**: Histogram of mass ratio from the preceding panel (projection of points onto the vertical axis), with Gaussian fit overlaid (and fit parameters shown).

cyan histogram in Figure 4.1 (**left**) and represents a naïve, but effective, realization of the mass distribution.

While the simple mass distribution utilizes the best distance estimate from the DPDF formalism, however, it discards much of the information contained in the posterior DPDFs. Given the asymmetric nature of many DPDFs, the most straightforward way to capitalize on this pool of information is through Monte Carlo realizations utilizing the MPDFs. Plotted in black in Figure 4.1 (**left**) is the mass distribution computed using the $M_{\rm ML}$ from the MPDFs. While there are noticeable deviations in the wings of the distributions, a two-sample KS test between these two distributions reveals that they are completely indistinguishable (to a probability better than machine precision). Despite the negligible differences between the distributions, the power of the MPDF-based mass distributions lies in the ability to create many Monte Carlo realizations of the mass distribution, randomly drawing a mass from each object's MPDF for each trial.

If the composite mass distributions between the two methods are indistinguishable, what of the masses of individual objects? The comparison of the mass derived from a simple application of Equation (4.2) with the $M_{\rm ML}$ drawn from the MPDF is shown in the middle panel of Figure 4.1,

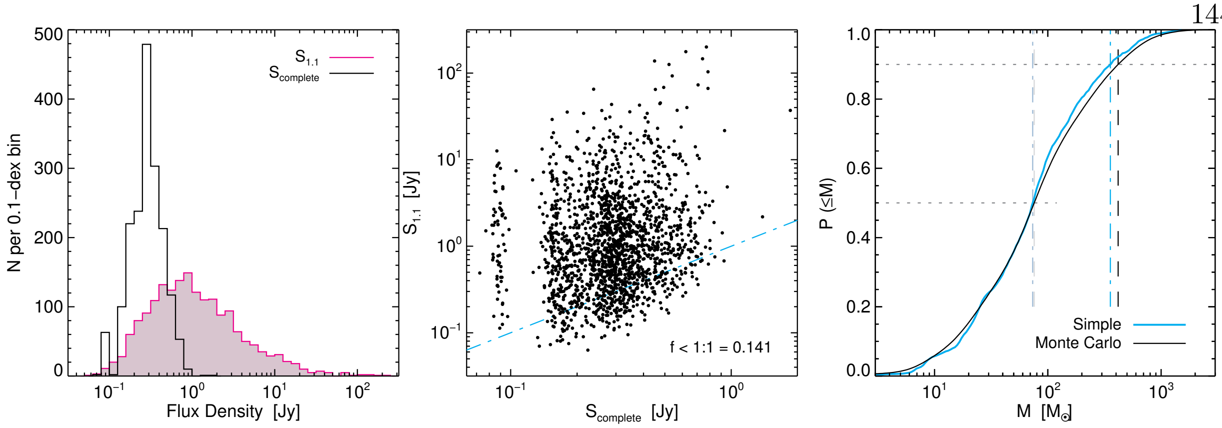

Figure 4.2 Computing mass completeness levels. **Left**: Histograms of the 40″-aperture flux density (this aperture is approximately one beam, magenta) and 5× rms noise (measured in Jy beam^{-1}, black) for each object in the Distance Catalog. **Middle**: Comparison of the flux density value and 5× rms noise for each source. The dot-dashed line marks the 1:1 relationship, with sources below this line being $< 5\sigma$ detections. **Right**: Cumulative distribution of the mass corresponding to 5× rms noise and the source distance from the posterior DPDF. The cyan curve represents the 'Simple' distribution, using the appropriate single distance estimator from the DPDF (see EB13). The black curve is the aggregate of 10^3 Monte Carlo realizations, with each realization randomly drawing a distance from each object's posterior DPDF and dust temperature from the lognormal distribution discussed in §4.4.1.1. Dotted horizontal lines identify 50% and 90% completeness. Light vertical dashed and dot-dashed lines identify the 50% completeness levels for both distributions, and the dark vertical lines mark the 90% completeness levels.

with the histogram of the mass ratio (vertical axis) shown in the right panel. The mean of the distribution is nearly unity, with a standard deviation of ≈ 3%, and differences between the masses are more significant for low-mass objects ($M \lesssim 100\,M_\odot$), with the $M_{\rm ML}$ becoming larger than the simple mass by up to 60%. The source of this trend is unclear, except that the fractional uncertainty in the flux density is larger for low-mass objects owing to their faintness.

4.3.2.2 Mass Completeness Function

(Sub-)Millimeter continuum surveys of thermal dust emission are flux-limited, so Malmquist bias must be addressed when analyzing the mass distribution of detected objects. It is relatively straightforward to identify flux-density completeness levels, but since more massive (and therefore brighter) objects are detectable through a larger Galactic volume than low-mass objects, it is imperative to define the mass criteria at which a survey is complete to some level. While small surveys of isolated regions at known distances (Ridge et al., 2006; Enoch et al., 2006; Rosolowsky et al., 2008a) can directly estimate strict mass completeness levels, heterogenous surveys of the entire Galactic plane must be content to have a more relaxed criterion.

We cast the issue of completeness levels in terms of a mass completeness **function**, which describes the fraction of objects in the survey area at any given mass that should be detected. Mass completeness begins with flux-density completeness, which is usually defined in terms of the rms noise in survey maps. For instance, the BGPS V1 catalog is 99% complete at 5× the rms noise flux density for any given field (Rosolowsky et al., 2010b), a value that is unchanged in version 2 (G13). Surveys that do not have uniform noise properties across the Galactic plane pose an additional complication to the process; the BGPS has a variable rms noise from field to field (see G13, their Figures 1 and 2). The flux-density completeness level must, therefore, be estimated on a field-by-field (or, better, source-by-source) basis. Estimates of rms noise in survey images, however, usually do not take the confusion limit into account, so these completeness levels should be viewed as lower limits.

The local flux-density completeness level for each catalog object is computed as 5× the mean of the BGPS noise maps within the object's catalog boundary (see Rosolowsky et al., 2010b). The minimum complete flux density for each object implies that any source over this brightness will be detected $\geq$ 99% of the time. To illustrate key aspects of this process, Figure 4.2 (**left**) shows the histogram of the completeness flux density for each source in black over the histogram of observed flux densities (red) for the Distance Catalog sources. The direct comparison of values for each source is shown in the middle panel, with the dot-dashed line marking the 1:1 relationship (observed flux density equals the completeness flux density). Approximately 14% of the objects in our sample have flux densities below the completeness value, meaning that they are $< 5\sigma$ detections. These objects represent a tip of the iceberg; similar objects exist in the observed fields but were not recovered.

Combining the completeness flux density with the object's distance information allows an estimate of the "minimum complete mass" for each object, meaning that any object at the specified $(\ell, b, d_\odot)$ more massive than this value will be detected $\geq$ 99% of the time. Repeating this process on a source-by-source basis allows for the effective marginalization over the variable noise properties across the Galactic plane. The mass completeness function is therefore the cumulative distribution of the minimum complete mass for all objects (Figure 4.2, **right**). The two curves in the plot correspond to using the single distance estimate from EB14 for each object (cyan), and an aggregation of 10^3 Monte Carlo realizations pulling masses from each object's MPDF (black). The curves are similar, but a two-sample KS test reveals they are not drawn from the same population at greater than 99.5% significance. Shown as horizontal dotted lines are the 50% and 90% completeness levels, and the vertical lines mark the intersections of the curves with these levels. The 50% completeness level is $\approx 75\,M_\odot$ for both curves, and the 90% completeness level is $\approx 400\,M_\odot$. The importance of these values is in limiting conclusions about the mass distributions (or other derived physical properties) to larger masses; extending conclusions to smaller-mass objects is not supported by the available data.

4.3.3 Fitting a Functional Form to the Mass Distribution

To compare the observed mass distribution for dense molecular cloud structures with theory and other observations, we fit both power law and lognormal functions to the data. The formulae and procedures for this fitting are discussed below, and are based on maximum-likelihood methods to avoid the pitfalls of working with binned data.

4.3.3.1 Power Law

The use of power-law functions for describing the mass distributions of stars and their precursors goes back to the original studies of the stellar initial mass function by Salpeter (1955). The mass function is defined as the number of objects either per logarithmic mass interval,

$$\xi(\log M) = \frac{dN}{d\log M} \propto M^{-x} \ , \tag{4.8}$$

or per linear mass interval,

$$\xi(M) = \frac{dN}{dM} = \frac{\xi(\log M)}{M \ \ln 10} = \left(\frac{1}{M \ln 10}\right) \frac{dN}{d\log M} \ , \tag{4.9}$$

where $\xi(M) \propto M^{-\alpha}$, and $x = \alpha - 1$ (Chabrier, 2003). Within this framework, the Salpeter power-law index is $x = 1.35$, $\alpha = 2.35$. For a detailed description of this parameterization and its consequences, see Swift & Beaumont (2010), Olmi et al. (2013), and references therein. The mathematical nature of a power law requires a finite lower limit on the mass range over which it is a valid descriptor of the data in addition to the observational constraint that the mass distribution turns over at $M \lesssim 1\,M_\odot$ (**cf.** Kroupa, 2001).

To fit a power-law function to the observed mass distribution, we use the maximum-likelihood method described by Clauset et al. (2009)[6] to estimate the power-law index and range over which it is valid. Briefly, assuming $\alpha > 1$ and $M_{\rm max} \gg M_{\rm min}$, the probability (or **likelihood**) that the observed masses (M_i) are drawn from a power-law distribution is proportional to

$$p_{\rm pl}\,(M\,|\,\alpha) = \prod_{i=1}^{n} \frac{\alpha - 1}{M_{\rm min}} \left(\frac{M_i}{M_{\rm min}} \right)^{-\alpha} , \tag{4.10}$$

where α is defined as in Equation (4.9). The maximum-likelihood value is computed by taking the derivative of the logarithm of the likelihood ($\mathcal{L}$)[7] with respect to α and finding the root (**i.e.**, $\partial\mathcal{L}/\partial\alpha = 0$). The power-law index estimator is

$$\hat{\alpha} = 1 + n \left[\sum_{i=1}^{n} \ln \frac{M_i}{M_{\rm min}} \right]^{-1} . \tag{4.11}$$

The minimum mass ($M_{\rm min}$) for which a power law is a good descriptor of the mass distribution is computed from the data. For each value in the mass distribution, Equation (4.11) is computed using that mass as $M_{\rm min}$. The cumulative distribution of $M \geq M_{\rm min}$ is compared with the analytic power law with index $\hat{\alpha}$, and the maximum difference between them is computed (**i.e.**, the K-S test D-statistic); the value of $M_{\rm min}$ that minimizes D is returned by the algorithm (see Clauset et al., 2009 for a complete description of the algorithm). The results presented here were computed using a `python` translation of the `PLFIT` routine from Clauset et al.[8]

4.3.3.2 Lognormal

[6] `http://tuvalu.santafe.edu/~aaronc/powerlaws/`

[7] The log-likelihood is used for ease of use, as the product in Equation (4.10) becomes a sum. While the absolute value of the function is changed by taking the logarithm, the location of the maximum is not.

[8] `https://github.com/keflavich/plfit`

Predicated on the properties of supersonic turbulence and the turnover in the IMF at $M \lesssim 1\,M_{\odot}$, the lognormal form of the mass function is also commonly fit to the mass distributions of dense molecular cloud structures (**cf.** Swift & Beaumont, 2010; Olmi et al., 2013). Analogous to the power-law case, we use a maximum-likelihood method for computing the best fit to the mass distribution. The likelihood of the observed masses being drawn from a lognormal distribution is proportional to

$$p_{\ln}\left(M \mid \mu, \sigma\right) = \prod_{i=1}^{n} \frac{C_{\ln}}{M} \exp\left[-\frac{(\ln M - \mu)^2}{2\sigma^2}\right] = \prod_{i=1}^{n} \frac{C_{\ln}}{M} \exp\left[-x^2\right] \ , \tag{4.12}$$

where the normalization factor

$$C_{\ln} = \sqrt{\frac{2}{\pi\sigma^2}}\ \left[\mathrm{erfc}(x_{\mathrm{min}}) - \mathrm{erfc}(x_{\mathrm{max}})\right]^{-1} \ , \tag{4.13}$$

and $\mathrm{erfc}(x)$ is the complimentary error function evaluated at x. The parameters $[\mu, \sigma]$ are the mean and width of the lognormal Gaussian, respectively, and $[x_{\mathrm{min}}, x_{\mathrm{max}}]$ are related to the minimum and maximum masses, respectively, over which the lognormal fit is valid. The maximum-likelihood estimators $[\hat{\mu}, \hat{\sigma}]$ are found by numerically solving $\nabla_{\mu,\sigma}\ \mathcal{L} = 0$ (**i.e.**, numerically maximizing $\mathcal{L}$), where

$$\begin{aligned} \mathcal{L} &= \ln\ p_{\ln}\left(M \mid \mu, \sigma\right) \\ &= n \ln C_{\ln}\ -\ \sum_{i=1}^{n} \ln\ M_i\ -\ \sum_{i=1}^{n} x_i^2\ . \end{aligned} \tag{4.14}$$

As with the power-law case, the limits $[M_{\mathrm{min}}, M_{\mathrm{max}}]$ over which the lognormal fit best describe the data must be estimated from the data. We employ the same K-S test D-statistic analysis as described above, but use a numerical minimizer to explore the 2-dimensional parameter space including M_{max}. Following the algorithmic structure of `PLFIT` (§4.3.3.1), we wrote a routine in `python` to maximize Equation (4.14) using Powell's Method inside a numerical minimization of the K-S test D-statistic using a downhill simplex optimization.[9]

[9] Both optimization steps utilized routines from the `scipy` library (`http://www.scipy.org`).

4.4 Results

4.4.1 Physical Properties of BGPS Sources

4.4.1.1 Catalog

We present a physical properties catalog for the 1,802 sources in the Distance Catalog. The mass, physical radius, and number density for each object were computed from the Monte-Carlo probability density functions as described in §4.3.1, where a dust temperature was drawn for each realization from a lognormal distribution with a mean of 20 K and full-width at half-maximum of 8 K (Battersby et al., 2011). The pdfs were constructed in log space to capture the spread in values over orders of magnitude, and error bars were computed to enclose 68.3% of the total probability, where the endpoints occur at equal probability (see the discussion of DPDF error bars in EB14). The error bars approximate the Gaussian $\pm 1\sigma$ region, but account for the asymmetric nature of the pdfs. The catalog of physical properties is shown in Table 4.1, along with relevant information from the BGPS V2.1 catalog (G13) and heliocentric distance (EB14).

Of particular note in the catalog are the 21% of sources that do not have a finite deconvolved angular radius (θ_R) in the BGPS V2.1 catalog (Equation 4.6). A brief analysis of these objects in the context of the flux-density distribution for all catalog sources reveals a strong trend whereby objects with non-physical θ_R are preferentially dim (Figure 4.3). The gray and blue histograms represent the "unresolved" sources in the full V2.1 catalog and the Distance Catalog, respectively, while the black and red histograms represent the resolved sources in the same groups.

The obvious correlation in the data points in the top panel (Spearman rank correlation coefficient $\rho = 0.753$) is a direct consequence of a universal scaling relationship in the dense interstellar medium first identified by Larson (1981) whereby the mass of a molecular cloud structure is roughly proportional to the square of its radius. Section 4.5.2 goes into more detail about this relationship, culminating in Equation (4.17), which is plotted as green lines in Figure 4.3 for $d_\odot = 1$ kpc (solid), 5 kpc (dotted), and 20 kpc (dashed).

Table 4.1. Physical Properties of BGPS V2 Sources

BGPS V2.1 Catalog Properties						Derived from the PDFs[a]		
Catalog Number	ℓ (°)	b (°)	$S_{\rm int}$[b] (Jy)	θ_R[c] ($''$)	$d_\odot$[d] (kpc)	M ($\log M_\odot$)	R (pc)	n ($\log$ cm^{-3})
2235	7.993	−0.268	6.23(0.49)	74.3	$12.16^{+0.66}_{-0.48}$	$4.09^{+0.12}_{-0.11}$	$4.39^{+0.24}_{-0.16}$	$2.80^{+0.12}_{-0.11}$
2254	8.187	0.482	0.61(0.15)	35.9	$2.02^{+0.86}_{-1.10}$	$1.58^{+0.38}_{-0.62}$	$0.41^{+0.11}_{-0.19}$	$3.53^{+0.34}_{-0.27}$
2256	8.207	0.190	1.44(0.22)	56.7	$2.80^{+0.82}_{-1.12}$	$2.20^{+0.28}_{-0.45}$	$0.82^{+0.17}_{-0.27}$	$3.20^{+0.24}_{-0.22}$
2261	8.249	0.180	2.29(0.29)	77.6	$3.36^{+0.58}_{-0.76}$	$2.52^{+0.21}_{-0.25}$	$1.29^{+0.20}_{-0.27}$	$2.90^{+0.18}_{-0.15}$
2262	8.263	0.168	2.82(0.32)	79.8	$2.94^{+0.70}_{-0.88}$	$2.52^{+0.23}_{-0.33}$	$1.19^{+0.23}_{-0.34}$	$3.01^{+0.19}_{-0.18}$
2265	8.281	0.164	1.00(0.15)	23.8	$3.26^{+0.60}_{-0.78}$	$2.14^{+0.21}_{-0.27}$	$0.39^{+0.06}_{-0.09}$	$4.09^{+0.18}_{-0.16}$
2295	8.545	−0.342	1.00(0.20)	35.5	$4.44^{+0.56}_{-0.74}$	$2.39^{+0.15}_{-0.18}$	$0.77^{+0.06}_{-0.08}$	$3.42^{+0.21}_{-0.22}$
2296	8.551	−0.296	0.25(0.10)	···	$12.38^{+0.48}_{-0.44}$	$2.74^{+0.19}_{-0.21}$	···	···
2297	8.579	−0.344	0.86(0.18)	37.3	$4.20^{+0.44}_{-0.54}$	$2.28^{+0.19}_{-0.19}$	$0.76^{+0.07}_{-0.09}$	$3.33^{+0.17}_{-0.17}$
2300	8.669	−0.406	0.44(0.11)	···	$11.92^{+0.48}_{-0.38}$	$2.92^{+0.16}_{-0.16}$	···	···

[a]See §4.3.1.

[b]The source-integrated $\lambda = 1.1$ mm flux density, uncertainties in parentheses.

[c]The deconvolved radius of the catalog object, computed using Equation (4.6).

[d]The appropriate distance estimate from EB14; $\overline{d}$ for objects near the tangent point and $d_{\rm ML}$ otherwise.

Note. — This table is available in its entirety in a machine-readable format in the online journal. A portion is shown here for guidance regarding its form and content.

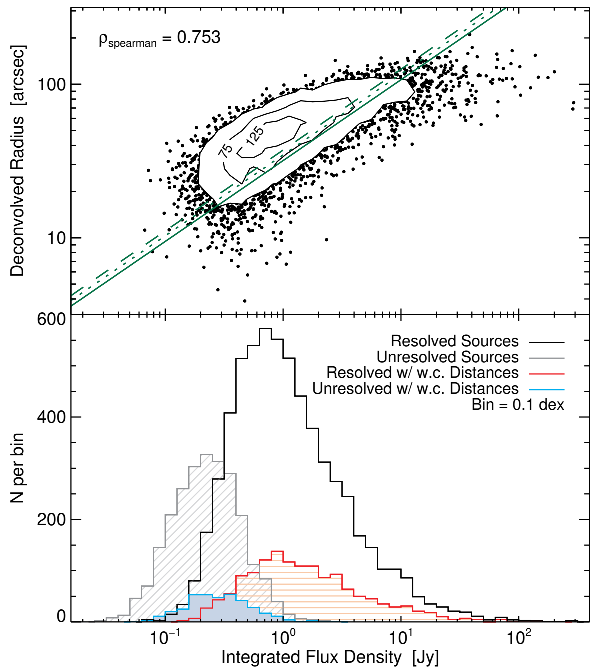

Figure 4.3 Brief analysis of resolved versus unresolved BGPS V2 sources. **Top**: Scatter plot of deconvolved radius from Bolocat versus integrated $\lambda = 1.1$ mm flux density. Evenly spaced contours represent the density of points per 0.1 dex $\times$ 0.1 dex bin. Only resolved sources are shown. Plotted in green are Equation (4.17) evaluated at $d_\odot = 1$ kpc (solid), 5 kpc (dotted), and 20 kpc (dashed) (see §4.5.2 for discussion). **Bottom**: Histograms of integrated $\lambda = 1.1$ mm flux density. Black represents all resolved sources (**i.e.**, the projection of the points in the top panel), and the gray diagonally hashed histogram represents the unresolved sources. Red horizontally hashed and blue solid histograms represent the resolved and unresolved populations, respectively, of the Distance Catalog (EB14).

The Larson relationships in the figure extend to low flux-density, but since the cataloging process truncates source boundaries where the flux density meets the local rms value (Rosolowsky et al., 2010b), source masks do not extend to the theoretical zero-flux-density isophot. The resulting catalog source may have an extent in the map smaller than the beam size, leading to a non-finite solution to Equation (4.6). To further aggravate the situation, the cataloger assumes that sources are spherical (or at least have circular projections on the sky), which is invalid for many of the fainter filamentary structures throughout the Galactic plane (**cf.** Molinari et al., 2010a). Future cataloging routines may be constructed that divide emission between filamentary and compact sources, and may decrease the number of BGPS sources with invalid deconvolved angular extent. For the present, however, unresolved sources are excluded from the following analyses that rely on

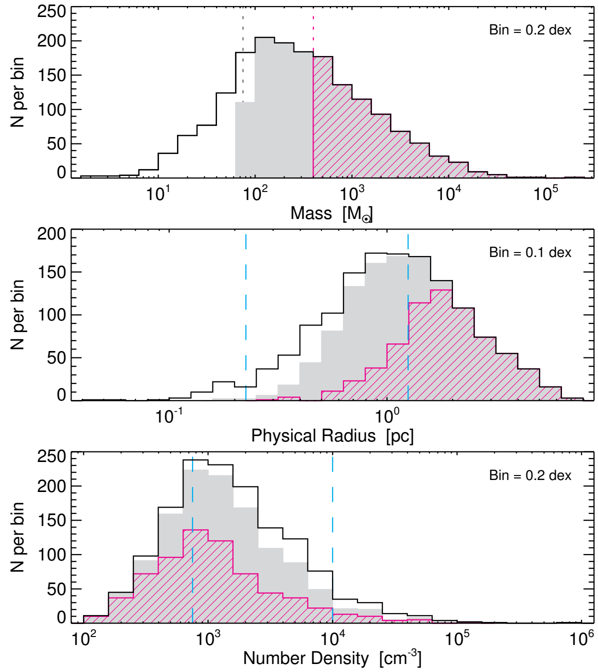

Figure 4.4 Physical property histograms for the Distance Catalog. Properties used are the maximum-likelihood values computed via the Monte Carlo method described in §4.3.1. The gray filled and magenta hashed histograms represent sources above the 50% and 90% mass completeness levels, respectively (see §4.3.2.2). **Top**: Total mass distribution. Vertical lines mark the locations of the 50% (gray) and 90% (magenta) completeness levels. **Middle**: Physical radius distribution. The vertical dashed lines identify the boundaries between cloud / clump / core as used by Dunham et al. (2011b). **Bottom**: Number density distribution; vertical lines as in the middle panel.

physical radius or number density.

4.4.1.2 Ensemble Physical Property Distributions

The ensemble physical property distributions from Table 4.1 are shown in Figure 4.4 to illustrate the range of BGPS objects int he Distance Catalog. This subsample is generally representative of the entire BGPS catalog, except that detection in a molecular transition line that traces dense gas (**e.g.**, HCO^+(3-2)) is strongly correlated with continuum flux density; dim sources are largely absent from the Distance Catalog (see EB14 for a complete discussion). In all three panels, sources above the 50% and 90% mass completeness levels are indicated by gray solid and magenta hashed histograms, respectively.

The mass distribution (top panel) peaks around 100 $M_\odot$, roughly coincident with that of the

$\ell = 30°$ Hi-GAL field analyzed by Olmi et al. (2013). Since over half of the Distance Catalog is in the range $20° \leq \ell \leq 40°$, this basic correspondence is encouraging. The completeness levels indicated by the filled histograms, however, imply that the location of the distribution peak is observationally biased; the predicted turnover at low mass is not constrained by the current observations.

Physical radius is illustrated in the middle panel of Figure 4.4, with vertical dashed lines marking plausible boundaries between molecular clouds (large R), clumps (intermediate R), and individual cores. Taking standard definitions from Bergin & Tafalla (2007), whereby clouds are $10^3 - 10^4\,M_\odot$ with sizes $2 - 15$ pc, clumps are $10^2 - 10^3\,M_\odot$ with sizes $0.3 - 3$ pc, and cores are $1 - 100\,M_\odot$ with sizes of about $0.03 - 0.2$ pc, we place approximate radius boundary markers at 1.25 pc between cloud and clump and 0.125 pc between clump and core, as in Dunham et al. (2011b, hereafter D11). These divisions, while rooted in physical distinctions based on virial ratio and capacity for Jeans fragmentation, are flexible and the boundaries are easily blurred by the continuum of structure in molecular cloud complexes (see §4.5.2), and by variations in Galactic environment.

Definitions notwithstanding, the typical size scale of BGPS-identified dense molecular cloud structures is clearly in the vicinity of 1-2 pc and may encompass multiple true but unresolved molecular cloud clumps. The mass completeness subsets illustrate the effects of Malmquist bias (coupled with the Larson relationship connecting mass and physical radius) whereby the 90% complete sample is nearly all cloud-scale ($R \gtrsim 1$ pc) objects, and the BGPS does not detect any large ($R \geq 2$ pc), lower-mass objects ($M \leq 400\,M_\odot$). This places a rough constraint on the minimum number density ($n \gtrsim 230$ cm^{-3}, Equation 4.7) detectable by the BGPS for large-scale objects.

The bottom panel of Figure 4.4 illustrates the number density distribution of resolved sources, which spreads across a wide dynamic range. Using the Bergin & Tafalla criteria along with Equation (4.7) yields approximate number density boundaries at $n = 750$ cm^{-3} between cloud and clump, and $n = 10^4$ cm^{-3} between clump and core (D11). These are shown as vertical dashed lines in Figure 4.4 (**bottom**). One striking feature of these histograms is the lack of complete

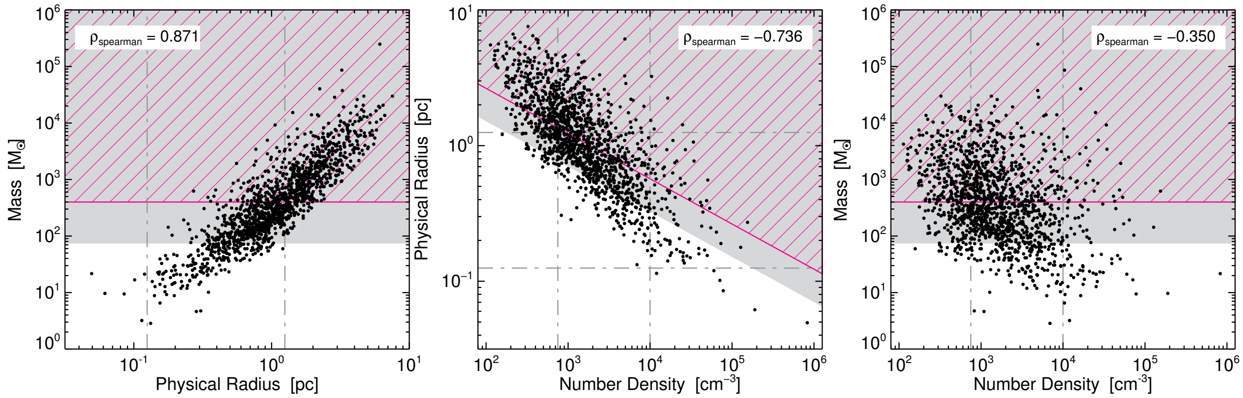

Figure 4.5 Inter-comparison of the physical properties for sources from the BGPS Distance Catalog with finite θ_R. In all panels, the gray region corresponds to the $M \geq 75\,M_\odot$ (50% mass completeness) region identified in Figure 4.4, and the magenta hashed region to $M \geq 400\,M_\odot$ (90% mass completeness). Spearman rank correlations are shown for each panel. **Left**: Mass versus physical radius. **Middle**: Physical radius versus number density. **Right**: Mass versus number density.

overlap between the completeness samples and the black histogram at the low-density end (that should correspond to the high-mass, large-radius, 90% complete sample), an overlap present in the R histograms in the middle panel. It is possible that non-uniform geometry of sources is causing the spread of the completeness samples over a large range, or that the algorithm by which the cataloging routine decomposes emission into discrete sources is doing so in a manner inconsistent with the underlying physical structure in molecular clouds.

An alternative visualization of the ensemble physical property distributions is presented in Figure 4.5, whereby the properties are plotted one against another to identify correlations and relationships. In all panels, the gray solid and magenta hashed regions again correspond to the 50% and 90% mass completeness levels. The left panel demonstrates the tight ($\rho = 0.871$) correlation between mass and physical radius suggested by the middle panel of Figure 4.4; this relationship is explored further in §4.4.4.

The middle panel of Figure 4.5 shows the inverse relationship between physical radius and number density that crosses constant-mass contours (the edges of the gray solid and magenta hashed regions). This relationship is not so tight as $M - R$ from the left panel, but still has a Spearman rank correlation coefficient $\rho = -0.736$. The right panel of the figure demonstrates the lack of localization of the 90% and 50% mass completeness histograms in the bottom panel of Figure 4.4. There is a less significant correlation ($\rho = -0.350$) between mass and number density, with some high-mass objects having an apparently high density as well, causing the spread of the 90% mass complete sample in the histogram in Figure 4.4 (**bottom**). One possible interpretation is that, despite the tight $M - R$ relationship, the continuum of structure within molecular cloud complexes is independent of number density. The other, more likely interpretation is that the assumed geometry of detected sources (§4.3.1.2) does not adequately describe the range of actual source geometry and that the cataloging algorithm decomposes emission into discrete sources inconsistent with the actual physical structure.

4.4.1.3 Physical Properties as a Function of Heliocentric and Galactocentric Position

A major challenge in characterizing the ensemble physical property distributions of detected molecular cloud structures is separating observational and data-processing systematic effects from underlying physical relationships. Systematic effects on the measurement of physical properties can lead to significant biases in the derived quantities. Because well-constrained distance estimates are the crucial pieces for understanding the physics of dense molecular cloud structures, it is instructive to analyze the functional dependences.

Shown in the left panels of Figure 4.6 are the mass, physical radius, and number density for sources plotted against heliocentric distance. Sources are plotted as gray dots, with the median value per 2-kpc bin marked by black squares. The dependence on heliocentric distance can be clearly seen in the median values as $M \sim d_\odot^2$, $R \sim d_\odot$, and $n \sim d_\odot^{-1}$, following Equations (4.2), (4.5), and (4.7), respectively. For the distribution of mass (top left panel), the lower hashed region identifies the mass associated with the mean 5σ flux density completeness level ($S_{\rm complete} = 0.30$ Jy, see Figure 4.2). Especially given the variable rms noise from field to field in the BGPS, sources may be detected in this region but completeness is not assured. The upper hashed region corresponds to the the flux density of the brightest BGPS source in the catalog; it is an empirical limit, plotted

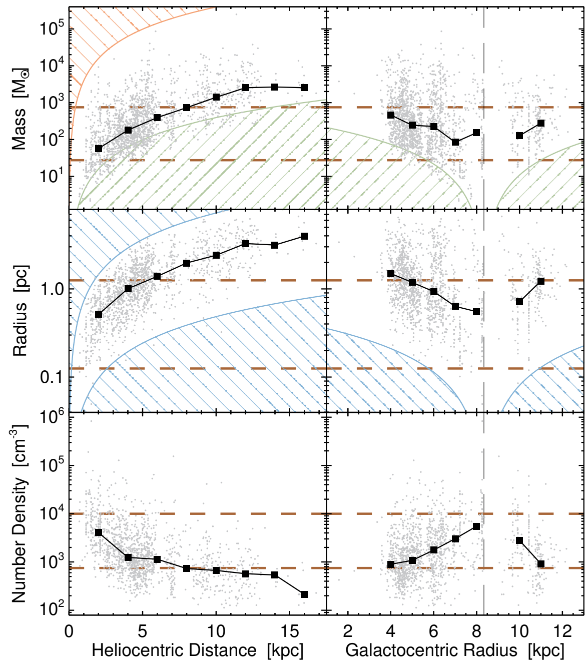

Figure 4.6 Physical properties as a function of heliocentric distance and Galactocentric radius. Grey dots mark individual objects from the BGPS distance catalog, and black squares identify the median values in 2-kpc ($d_\odot$) or 1-kpc ($R_{\rm gal}$) bins. Horizontal dashed lines indicate the rough boundaries between clouds, clumps, and cores. **Top Row**: Computed mass for each object. The lower, green hashed region represents the mean 5σ flux-density completeness level as a function of heliocentric distance. For the right panel, this region is computed for $R_{\rm gal}$ along the $\ell = 0°, 180°$ line. **Middle Row**: Physical radius. The lower hashed region corresponds to $\theta_R = 10''$, while the upper marks $\theta = 150''$. For the right panel, the lower region is again computed for $R_{\rm gal}$ along $\ell = 0°, 180°$. **Bottom Row**: Number density.

to illustrate the regions of parameter space devoid of observed sources. The large collection of sources around $d_\odot = 5$ kpc has a median mass $\approx 300\,M_\odot$, whereas the median grows to $\approx 2000\,M_\odot$ for objects $\gtrsim 10$ kpc. The physical radius plot (middle left panel) shows similar effects, with the hashed regions marking $\theta_R \leq 10''$ (typical minimum angular radius) and $\theta_R \geq 150''$ (approximate maximum recoverable radius; see G13). The horizontal dashed lines represent the $R = 0.125$ pc and $R = 1.25$ pc divisions between physical object categories discussed above. Finally, the number density distribution (bottom left panel) displays the expected trend, except that the median values remain somewhat constant between 500 cm$^{-3} \leq n \leq 1000$ cm^{-3} over a large range of heliocentric distance. Since this flattening is not observed for physical radius (which also depends linearly on $d_\odot$), this range of n may represent a typical value for dense molecular cloud structures, subject to the geometric and cataloging caveats discussed above.

The right panels in Figure 4.6 illustrate the dependence of these physical properties on Galactocentric radius. The vertical dashed line identifies $R_{\rm gal} = R_0 = 8.34$ kpc (Reid et al., 2014) for reference. Black squares mark the median values in 1-kpc bins, and roughly show the effects of heliocentric distance (**i.e.**, objects far from $R_{\rm gal} = R_0$ are at large $d_\odot$). Gaps in the black function represent bins with $N \leq 20$ sources, where the median value is skewed by outliers. The gap in sources (gray dots) at $R_{\rm gal} = 8.5 - 9.5$ kpc is pointed out in EB14 as arising from a unique combination of BGPS coverage region and a kinematic avoidance zone (where kinematic distances are unreliable in the face of non-circular motions about the Galactic center), and is not likely a true feature in the underlying Galactic distribution of dense molecular gas. Many of the dense molecular cloud structures observed in the Molecular Ring / Scutum-Centarus Arm feature at $R_{\rm gal} = 4 - 5$ kpc, $d_\odot = 3 - 6$ kpc, correspond the physical properties typical of molecular cloud clumps.

The systematic effects of heliocentric distance will obscure some of the underlying physical trends in properties as a function of Galactocentric radius. For instance, the spike in median number density near $R_{\rm gal} = R_0$ is due to the population of nearby ($d_\odot \lesssim 2$ kpc) core-scale objects. The observed decrease in mass for dense molecular cloud structures from $R_{\rm gal} = 4$ kpc to $R_{\rm gal} =$ 7 kpc is consistent with the trend observed for GMCs by Roman-Duval et al. (2010).

4.4.2 Mass Function Fits for Dense Molecular Cloud Structures

4.4.2.1 Mass Distribution Trials

To employ the full information available from the MPDFs (and ultimately the DPDFs from EB14), we utilized a large number of Monte Carlo trials to fit functional forms to the mass distribution of dense molecular cloud structures. For each trial, we randomly drew a single value from the MPDF of each object (recall that the MPDFs were constructed via Equation (4.2) by drawing values from the object's DPDF and flux density distribution, and selecting a temperature from a

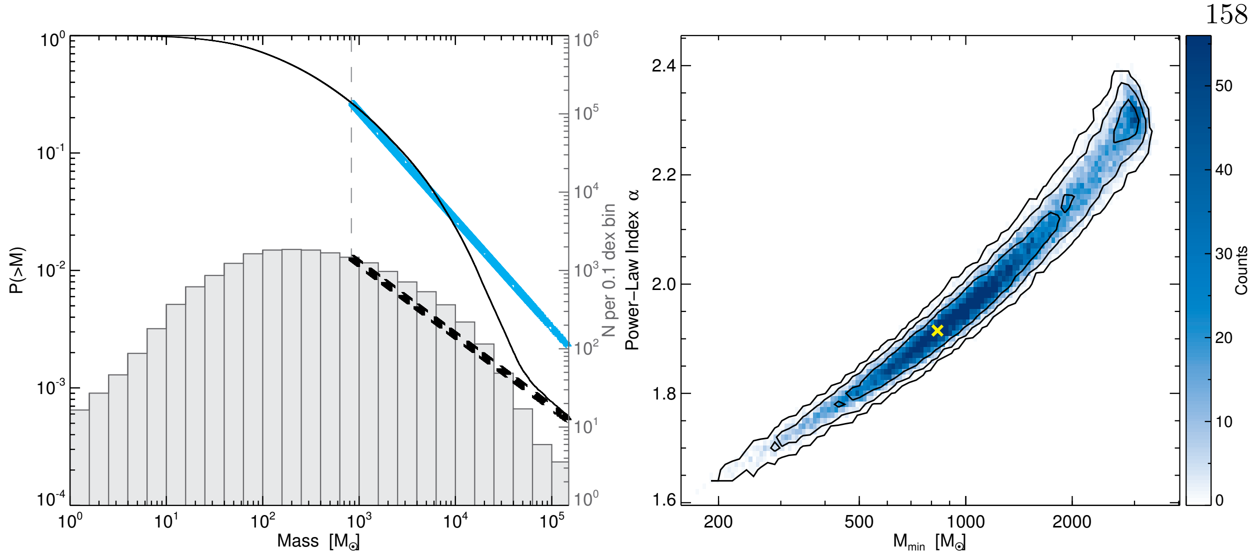

Figure 4.7 Power-law fit to the aggregate mass distribution. **Left**: The complementary cumulative distribution function (solid black curve) and differential mass distribution (gray histogram) for the Distance Catalog. This distribution was constructed from 10^3 Monte Carlo draws from the MPDFs for each object, and normalized to a sum of $N = 1802$ to create a an average mass distribution that marginalizes over the uncertainties in the mass of each object. The solid cyan and dashed black curves represent the maximum-likelihood power-law index (from right panel), and the vertical gray dashed line marks the maximum-likelihood minimum mass for which a power law describes the data. **Right**: The two-dimensional histogram of the power-law fit parameters for each of 20,000 Monte Carlo trials, whereby a power law was fit to a single realization of the mass distribution. Contours enclose 68.3% (highest), 95.4%, and 99.7% of the points in the plane. The yellow X marks the largest concentration of points, or the joint maximum-likelihood values of $\hat{\alpha}$ and $M_{\rm min}$.

universal temperature distribution) to create an independent realization of the mass distribution. Each realization was fit with both a power law (§4.3.3.1) and lognormal (§4.3.3.2) function, and the fit parameters recorded. This process was repeated $N = 20,000$ times, a value chosen to balance appropriately sampling the parameter spaces for the functional fits with computation time. The collected fit parameters from the Monte Carlo process were then plotted as two-dimensional histograms to identify joint confidence intervals. Below, we discuss the results of the power-law (§4.4.2.2) and lognormal (§4.4.2.3) fits to the entire data set, and in §4.4.3 we divide the data set into astrophysically meaningful subsets to identify mass functions of interest. Maximum-likelihood fit parameter values for all fits are listed in Table 4.2.

4.4.2.2 Power-Law Fit

The power-law fits to the complete mass distribution are shown in Figure 4.7. The left panel illustrates the mass distribution both as the complementary cumulative distribution function (CCDF, solid black)[10] and as the differential mass distribution (gray histogram). Both distributions represent an aggregate distribution of 10^3 Monte Carlo realizations of a single mass distribution, as described above, normalized so the sum equals 1,802 objects. This aggregate is independent of the $N = 20,000$ Monte Carlo fit trials, and smoothes the effects of small sample size and utilizes the full amount of information encoded in the MPDFs. The power-law fits (derived from the data in the right panel) are shown for the CCDF (cyan line) and the differential mass distribution (black dashed line).

The power-law fitting algorithm (`PLFIT`) returns two parameters, the power-law index $\hat{\alpha}$ and the minimum mass over which the data may be described by a power law ($M_{\rm min}$). The two-dimensional histogram of the 20,000 parameter pairs from the Monte Carlo trials are shown in the right panel of Figure 4.7. The color scale indicates the number of points in each pixel of the image, and contours enclose 68.3% (highest contour), 95.4%, and 99.7% of the parameter pairs, meant to illustrate joint confidence regions. The high degree of correlation between $\hat{\alpha}$ and $M_{\rm min}$ is indicative of the continuously curved nature of the mass distribution shown in the left panel (where a power law is a straight line). The yellow X in the right panel marks the peak of the two-dimensional histogram, or the joint maximum-likelihood parameter values ($\hat{\alpha} = 1.92,\ M_{\rm min} = 831\,M_\odot$). This minimum mass is above the 90% completeness level described in §4.3.2.2 ($\approx 400\,M_\odot$), meaning the power-law fit is not affected by Malmquist bias. Note the secondary peak of points near $\hat{\alpha} = 2.3,\ M_{\rm min} = 3000\,M_\odot$ in the right panel; there appears to be a sharp steepening of the mass distribution at the high-mass end (see §4.5.3.2 for discussion). The distribution truncates around $M_{\rm min} = 3500\,M_\odot$ due to a requirement in the fitting algorithm for a minimum number of data points required for a robust fit (see Clauset et al., 2009). This truncation is a general feature of

[10] The CCDF is simply 1 - CDF, where CDF is the typical cumulative distribution function. The CCDF measures the probability of finding an object with mass **greater** than a given value, whereas the CDF describes the probability of finding an object with mass **less** than a given value. The CCDF is used to illustrate the distribution in a mmer robust against fluctuations caused by finite sample size.

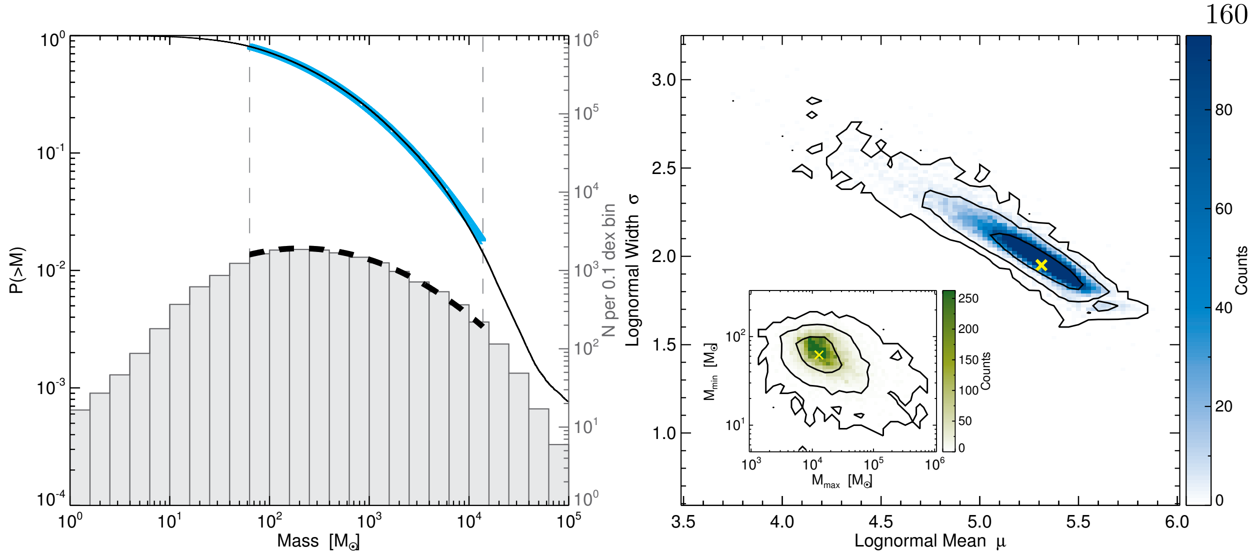

Figure 4.8 Lognormal fit to the aggregate mass distribution. **Left**: The complementary cumulative distribution function (solid black curve) and differential mass distribution (gray histogram) are identical to Figure 4.7. The blue and black dashed curves represent the maximum-likelihood lognormal fit (from the right panel), and the vertical gray dashed lines mark the maximum-likelihood minimum and maximum mass for which a lognormal describes the data. **Right**: The two-dimensional histogram of the lognormal fit parameters $[\mu, \sigma]$ for each of 20,000 Monte Carlo realizations of the mass distribution. Contours enclose 68.3% (highest), 95.4%, and 99.7% of the points in the plane. The yellow X marks the largest concentration of points, or the joint maximum-likelihood values. The inset shows the parameter space distribution for the pair $[M_{\rm min}, M_{\rm max}]$.

the fits in §4.4.3.

4.4.2.3 Lognormal Fit

The continuously curved nature of the mass CCDF makes fitting or interpreting a power law function for the entire data set difficult at best. Furthermore, distance inhomogeneities in large-scale surveys such as the BGPS drive the composite observed mass distribution towards a lognormal (see below). Applying this functional form results in the fits shown in Figure 4.8. The CCDF and differential mass distributions in the left panel are identical to Figure 4.7 (**left**), with the curves once again being drawn from the right panel.

Whereas the power-law fit returns two parameters, the lognormal fitting optimizes the functional fit over four parameters: the lognormal characteristic $[\mu, \sigma]$, and also both end of the mass range $[M_{\rm min}, M_{\rm max}]$ over which the lognormal fit is valid. This yields six planes into which the parameters may be sliced; two are shown in the right panel of Figure 4.8. Perhaps the most fundamental is the $[\mu, \sigma]$ plane, shown in the background, since these define the position and shape of the function. As with the power-law fit, shown here is the two-dimensional histogram of the 20,000 Monte-Carlo trial fit parameters, with contours representing the same enclosed fractions of points. Both parameters are tightly constrained, with $\hat{\mu} = 5.30 \pm 0.15$ ($170\,M_\odot \leq M_\mu \leq 230\,M_\odot$)[11] and $\hat{\sigma} = 1.96^{+0.12}_{-0.08}$, where the uncertainties represent the marginalized 68.3% confidence intervals. The joint maximum-likelihood values (marginalized over $[M_{\rm min}, M_{\rm max}]$) occur at $\mu = 5.32,\ \sigma = 1.96$, and are reflected in the solid cyan and dashed black curves in the left panel. Since the peak occurs at $M_\mu = e^{5.32} = 204\,M_\odot$, or below the $400\,M_\odot$ 90% completeness level, it may be somewhat biased and should be viewed as an upper limit.

Inset in the right panel is the two-dimensional histogram of the $[M_{\rm min}, M_{\rm max}]$ plane, showing the relationship between the bounding parameters. There is scant correlation, except that the spread along the diagonal from top-left to bottom-right represents contraction and expansion about a central value. The optimum values of these parameters (marginalized over $[\mu, \sigma]$) are indicated by the yellow X in the inset ($M_{\rm min} = 63\,M_\odot,\ M_{\rm max} = 13800\,M_\odot$), and are marked in the left panel by vertical gray dashed lines. This minimum mass extends down to the 50% mass completeness level ($\approx 75\,M_\odot$), and its location may not be entirely reliable, as the peak of the fitted lognormal function always falls below the 90% completeness level ($\mu < 6.0 = \ln\, 400\,M_\odot$). While there is uncertainty in the exact location of the turnover, the data here provide strong evidence for the existence of a turnover at low masses. The other four parameter combinations do not yield significant correlation or additional information about the fit, other than to illustrate that a tight mass range ($M_{\rm min}$ very near $M_{\rm max}$) leads to ill-constrained values for $[\mu, \sigma]$.

Comparison of the lognormal fit in Figure 4.8 with the power-law fit in Figure 4.7 yields some interesting notions. As expected, the minimum mass for the power-law fit is well above the peak of the lognormal distribution. In fact, the 99.7% contour in Figure 4.7 (**right**) terminates at

[11] The lognormal fits are done as a function of $(\ln M)$, so the conversion to linear mass requires the natural exponential base rather than base 10.

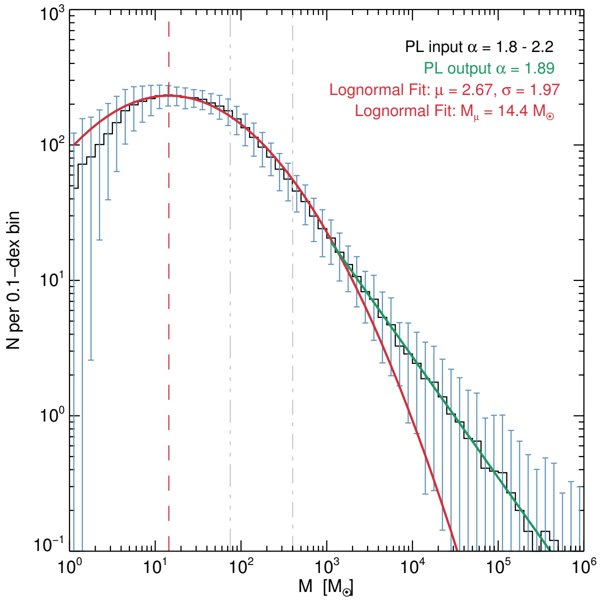

Figure 4.9 Simulated aggregate mass distributions of star-forming regions (SRFs) at a variety of distances. Only sources in those regions that would be detected by the BGPS (**i.e.**, $S \geq 3\sigma$) are included. **Left**: Composite of 100 SFRs, each with a power-law mass distribution with index $1.8 \leq \alpha \leq 2.2$. The error bars represent the spread among 100 Monte Carlo realizations. The green line represents the power-law fit to the high-mass end of the distribution, and the red curve is the lognormal fit to the peak. Vertical dot-dashed lines represent the 50% ($M = 75\,M_\odot$) and 90% ($M = 400\,M_\odot$) mass completeness levels, and the vertical dashed red line marks the M_μ of the lognormal fit. **Right**: Composite of 100 SFRs, each with a lognormal mass distribution with mean $4.5 \leq \mu \leq 6.5$ and width $1.5 \leq \sigma \leq 2.5$. Other features as at left.

about this value. Furthermore, the continuous curvature of the mass distribution is well-matched to a lognormal function over nearly 2.5 orders of magnitude. To place this continuous curvature in context, Table 4.3 lists the power-law index of a line tangent to the lognormal curve in Figure 4.8 (**left**) at a range of masses. These are significantly shallower than the corresponding contours in Figure 4.7 (**right**), but represent tangential power laws that do not attempt to follow the data; they are presented to illustrate the continuous change in the power-law index as a function of minimum mass for lognormal-type mass distributions.

4.4.2.4 The Inevitability of a Lognormal + Power-Law Mass Function?

On an observational note, the aggregate mass distribution, as shown in Figure 4.8 (**left**), should be expected to be well fit by a lognormal function with a power-law tail regardless of the

Table 4.2. Mass Distribution Functional Fit Parameter Values

Cut Type	Subset[a] Name	N	Power-Law Fit $\hat{\alpha}$	$M_{\rm min}$ (log $M_\odot$)	Lognormal Fit $\hat{\mu}$ (ln $M_\odot$)	$\hat{\sigma}$ (ln $M_\odot$)	$M_{\rm min}$[b] (log $M_\odot$)	$M_{\rm max}$[b] (log $M_\odot$)
	Distance Catalog	1802	$1.94^{+0.16}_{-0.12}$	$3.02^{+0.44}_{-0.11}$	$5.30^{+0.15}_{-0.15}$	$1.96^{+0.12}_{-0.08}$	$1.84^{+0.08}_{-0.16}$	$4.08^{+0.24}_{-0.12}$
$d_\odot$	Nearby	1390	$2.04^{+0.09}_{-0.09}$	$2.83^{+0.18}_{-0.17}$	$4.93^{+0.10}_{-0.12}$	$1.72^{+0.06}_{-0.02}$	$1.84^{+0.12}_{-0.12}$	$4.14^{+0.18}_{-0.24}$
	Distant	412	$1.91^{+0.21}_{-0.05}$	$3.10^{+0.25}_{-0.07}$	$7.00^{+0.08}_{-0.05}$	$1.50^{+0.12}_{-0.06}$	$1.84^{+0.16}_{-0.08}$	$4.14^{+0.24}_{-0.18}$
$R_{\rm gal}$	Sct-Cen Arm	930	$1.97^{+0.08}_{-0.09}$	$2.96^{+0.12}_{-0.19}$	$5.43^{+0.12}_{-0.23}$	$1.70^{+0.24}_{-0.02}$	$1.84^{+0.12}_{-0.08}$	$3.96^{+0.24}_{-0.18}$
	Sgr / Local Arms	638	$2.03^{+0.01}_{-0.27}$	$3.23^{+0.03}_{-0.41}$	$4.98^{+0.72}_{-0.85}$	$2.74^{+0.70}_{-0.52}$	$1.84^{+0.12}_{-0.12}$	$3.96^{+0.18}_{-0.18}$
	Per / Outer Arms	234	$1.94^{+0.04}_{-0.09}$	$2.58^{+0.04}_{-0.13}$	$5.55^{+0.20}_{-0.18}$	$1.70^{+0.04}_{-0.24}$	$1.84^{+0.16}_{-0.12}$	$3.96^{+0.30}_{-0.18}$
R	Clump / Core	819	$1.94^{+0.10}_{-0.11}$	$3.03^{+0.16}_{-0.20}$	$5.12^{+0.25}_{-0.45}$	$2.16^{+0.30}_{-0.18}$	$1.84^{+0.12}_{-0.12}$	$4.14^{+0.18}_{-0.24}$
	Cloud	608	$1.87^{+0.07}_{-0.23}$	$2.90^{+0.13}_{-0.57}$	$5.45^{+0.18}_{-0.42}$	$2.00^{+0.26}_{-0.14}$	$1.84^{+0.12}_{-0.12}$	$4.14^{+0.24}_{-0.18}$
n	Clump / Core	1021	$1.89^{+0.08}_{-0.23}$	$2.95^{+0.20}_{-0.60}$	$4.90^{+0.30}_{-0.62}$	$2.26^{+0.36}_{-0.18}$	$1.84^{+0.12}_{-0.08}$	$4.14^{+0.24}_{-0.18}$
	Cloud	406	$2.00^{+0.04}_{-0.17}$	$3.03^{+0.04}_{-0.29}$	$5.80^{+0.12}_{-0.23}$	$1.74^{+0.24}_{-0.04}$	$1.84^{+0.12}_{-0.16}$	$4.14^{+0.24}_{-0.24}$
Other	Blind ($\ell \leq 90°$)	1605	$1.92^{+0.36}_{-0.10}$	$2.98^{+0.47}_{-0.12}$	$5.35^{+0.17}_{-0.18}$	$2.04^{+0.14}_{-0.12}$	$1.84^{+0.12}_{-0.12}$	$4.18^{+0.18}_{-0.24}$
	$R \geq 0.2$ pc	1389	$1.91^{+0.11}_{-0.09}$	$2.99^{+0.18}_{-0.18}$	$5.20^{+0.20}_{-0.27}$	$2.10^{+0.24}_{-0.10}$	$1.84^{+0.12}_{-0.12}$	$4.18^{+0.18}_{-0.24}$
	Mixed Protocluster	637	$1.92^{+0.13}_{-0.06}$	$2.86^{+0.17}_{-0.19}$	$6.45^{+0.10}_{-0.12}$	$0.68^{+1.02}_{-0.10}$	$2.12^{+0.32}_{-0.08}$	$3.16^{+0.12}_{-0.18}$
	$M \geq 2000\, M_\odot$	231	$2.24^{+0.03}_{-0.07}$	$3.47^{+0.03}_{-0.05}$	$8.38^{+0.05}_{-0.05}$	$0.80^{+0.10}_{-0.08}$	$1.84^{+0.04}_{-0.04}$	$4.30^{+0.30}_{-0.24}$

[a]See §4.4.3 for definitions of these subsets.

[b]The present implementation of the lognormal fitter tends to find mass bounds very near the inital guess. This behavior is being investigated, but initial results indicate that the exact bounds have little impact on the $[\hat{\mu}, \hat{\sigma}]$ returned.

Table 4.3. Power-Law Index for Lognormal Function

Mass $M_\odot$	α
2×10^2	0.99
5×10^2	1.23
1×10^3	1.41
2×10^3	1.59
5×10^3	1.83
1×10^4	2.01
2×10^4	2.19
5×10^4	2.43

actual underlying distribution of source mass. Given the finite angular resolution and flux-limited nature of the BGPS (and similar surveys), it is possible to simulate the observed mass distribution of an inhomogenous sample spread across the Galactic plane.

We conducted a series of Monte Carlo trials, whereby we "observed" star-forming regions (SFRs) whose distances were drawn from the $d_\odot$ distribution of the Distance Catalog. Each SFR consisted of 100 objects drawn purely from either a power-law (Trial 1) or lognormal (Trial 2) mass distribution. A physical radius was assigned using Equation (4.15) (see §4.4.4) to estimate whether any given object would be resolved by the BGPS, and the extent to which its flux density (computed from the mass) would be diluted in the beam. Using an average BGPS noise level, sources were culled from the list that would not be detected ($S < 3\sigma$). The aggregate distribution of "detected" sources was recorded, and the process repeated for $N = 100$ Monte Carlo realizations.

The left panel of Figure 4.9 shows the results of Trial 1, where the underlying power-law mass distribution was allowed to range over $1.8 \leq \alpha \leq 2.2$. The aggregate shows a strong power-law tail with $\alpha = 1.9$, and a lognormal distribution at lower mass, peaking at $M_\mu \approx 15\,M_\odot$. While the location of the lognormal peak is well below the completeness levels of the BGPS data set, these simulations do not account for real features in the data, such as source confusion and hierarchical (fractal) structure, and likely overestimate the number of low-mass sources visible. Trial 2 (right panel) constructed individual lognormal mass distributions using the ranges $4.5 \leq \mu \leq 6.5$ and $1.5 \leq \sigma \leq 2.5$. The resulting aggregate mass distribution shows a steeper ($\alpha \approx 2.3$) power-law tail at high mass, but has a lognormal fit much closer to that of the full BGPS set in Figure 4.8.

These simulations are not meant as definitive proof of the underlying mass distributions, but rather to demonstrate the inevitability of a lognormal distribution with a power-law tail at high mass. While supersonic turbulence in molecular clouds does lead to a lognormal density distribution, Tassis et al. (2010) point out that other processes also produce this form. The principal issue illustrated by Figure 4.9, however, is that distance inhomogeneities lead to a lognormal + power-law mass distribution regardless of the underlying source distribution. In the next section, we explore the effects of trying to construct more homogenous subsets of the data on the resulting

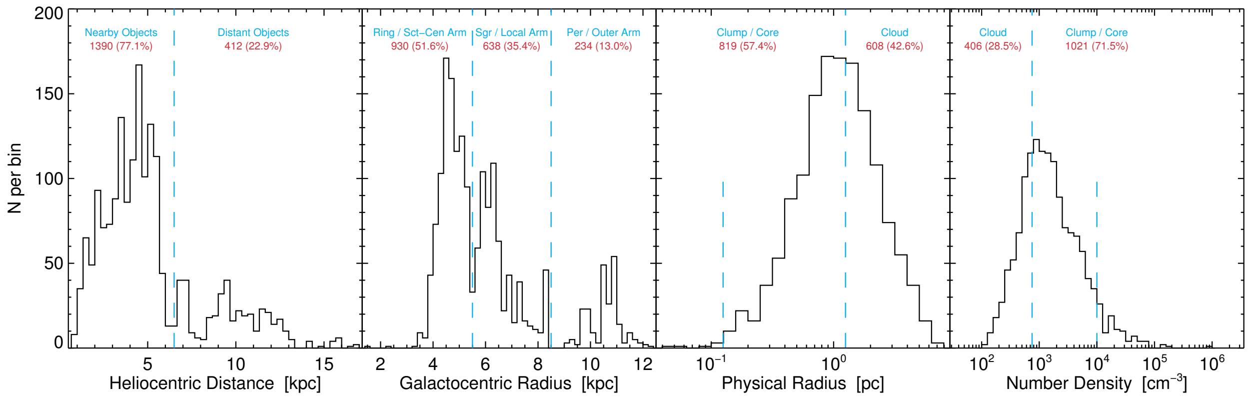

Figure 4.10 Astrophysically cut subsets. Shown are the distributions of four physical quantities by which the collection of sources may be divided for further study. The vertical dashed lines in each panel indicate the division points, and the text identifies the subsets and also the number of objects assigned to that group. For the two rightmost panels, only sources with finite θ_R are shown (see §4.4.1.1).

mass function fits.

4.4.3 Mass Functions Based on Astrophysical Cuts

Distance Catalog objects have wide-ranging distributions of heliocentric distance, Galactocentric radius, physical radius, and number density. These wide distributions suggest that the present sample is a rather heterogenous set of objects; conclusions drawn from their aggregate mass distribution may be affected by global changes in physical conditions and properties and not necessarily representative of any one type of physical object. Indeed, we know the BGPS generally identifies cores to a distance of 1 kpc, clumps from 1 kpc to 7 kpc, and clouds beyond 7 kpc (D11). As such, it is instructive to divide the total collection of sources into astrophysically meaningful subsets and analyze the resulting mass function fits individually.

The physical quantities chosen to discriminate objects for further analysis are shown in Figure 4.10. Each panel of the figure shows the distribution of sources with dashed vertical lines marking the boundaries between the subsets. The text in each panel identifies the physically significant subset and indicates the fraction of sources contained therein. Note that in the physical radius and number density panels, there is a third region for molecular cloud cores, but they are

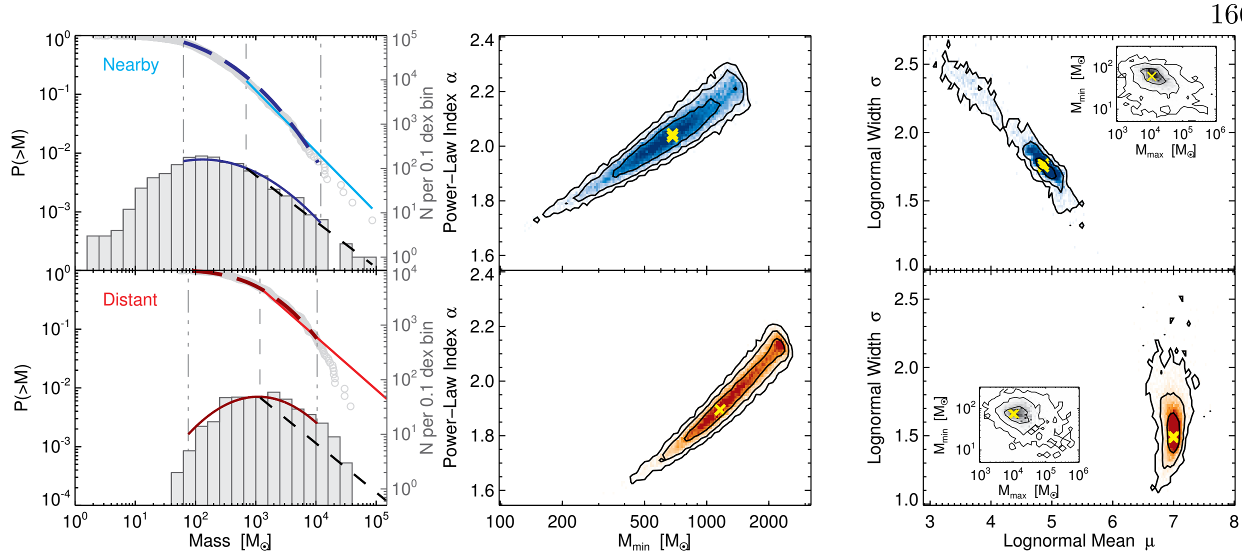

Figure 4.11 Mass functions of subsets in heliocentric distance. Top row illustrates the nearby subset ($d_\odot \leq 6.5$ kpc; lines in blue) and the bottom row more distant objects (lines in red). **Left**: The mass distribution shown as the CCDF (gray circles) and differential mass distribution (gray histogram). The power-law fits are shown as solid light-colored (CCDF) and dashed black (histogram) lines, and the lognormal fits as long-dashed (CCDF) and solid (histogram) dark-colored lines. **Middle**: Joint confidence intervals for the power-law fit, as in Figure 4.7. **Right**: Joint confidence intervals for the lognormal fit, as in Figure 4.8.

considered with the clump-scale objects to avoid biases from small-number statistics. In addition to discussing subsets based on each of these quantities below, we also present some heterogenous cuts in search of a "protocluster" mass function (§4.4.3.4). The implications of these subsets and their functional fits are discussed in §4.5.3.

4.4.3.1 Heliocentric Distance

The heliocentric distance histogram for Distance Catalog objects (Figure 4.10, **far left**) exhibits a minimum at $d_\odot \approx 6.5$ kpc. Beam-scale BGPS objects at this distance have $R \approx 1$ pc, indicating that most of the objects seen beyond this point are more likely to be on the scale of GMCs than molecular cloud clumps. Dividing the sample here creates a "nearby" group containing 77% of the objects, with the remaining 23% in the "distant" group. Figure 4.11 illustrates the mass distributions for both sets, as well as the functional fit parameter confidence intervals in the manner of Figures 4.7 and 4.8. The left column shows the mass distributions with both power-law and lognormal fits plotted. The middle column displays the joint confidence intervals for the $[\hat{\alpha}, M_{\rm min}]$ power-law parameters, and the right column the $[\hat{\mu}, \hat{\sigma}]$ and $[M_{\rm min}, M_{\rm max}]$ lognormal parameter joint confidence intervals. For both the middle and right columns, the rows are have the same plotting ranges for easy visualization of the differences between the fits to the subsets.

We begin with the nearby sample (top row). Not polluted with the more distant objects, the distribution strongly resembles a power law with $\hat{\alpha} = 2.04 \pm 0.09$ for $M \gtrsim 700\,M_\odot$, as evidenced by the quasi-symmetric joint confidence intervals in the middle panel. The power-law fit is shown in the left panel atop the mass distributions as a cyan line (for the CCDF) and black dashed line (for the differential mass distribution). The lognormal fit (right panel) shows a strong preference for values near the joint maximum-likelihood $(\hat{\mu}, \hat{\sigma}) = (4.88, 1.76)$. The low-probability tail towards the top-right corner indicates the need for a wider gaussian to fit the data if the mean is moved to lower mass.

For comparison, the distant sample (bottom row) does not exhibit a strong power-law shape, as the middle panel suggests a continuum of $[\hat{\alpha}, M_{\rm min}]$ across the 20,000 Monte Carlo trials, with a marginalized maximum-likelihood power law index $\hat{\alpha} = 1.91^{+0.12}_{-0.05}$. The small sample size of the distant group also means that each independent realization of the mass distribution will demonstrate more variability in the fit parameters than for a large-N sample. Despite this potentially causing volatility in the power-law fits, however, the lognormal fits have a tightly constrained mean around $M_\mu = e^{7.00} = 1100\,M_\odot$. The significant difference between the $\hat{\mu}$ for the nearby and distant subsets clearly illustrates the systematic $M \sim d_\odot^2$, indicating, perhaps, a real division between physical categories of objects. Under this assumption, the nearby clump-scale objects have a power-law index near $\hat{\alpha} = 2.0$, while the more distant cloud-scale objects have $\hat{\alpha} \approx 1.9$ (although with a long tail to higher α), hinting at the possibility of a power-law progression from GMCs to the stellar IMF.

4.4.3.2 Galactocentric Radius

Galactic environment can influence the properties and evolution of star-forming regions (Ken-

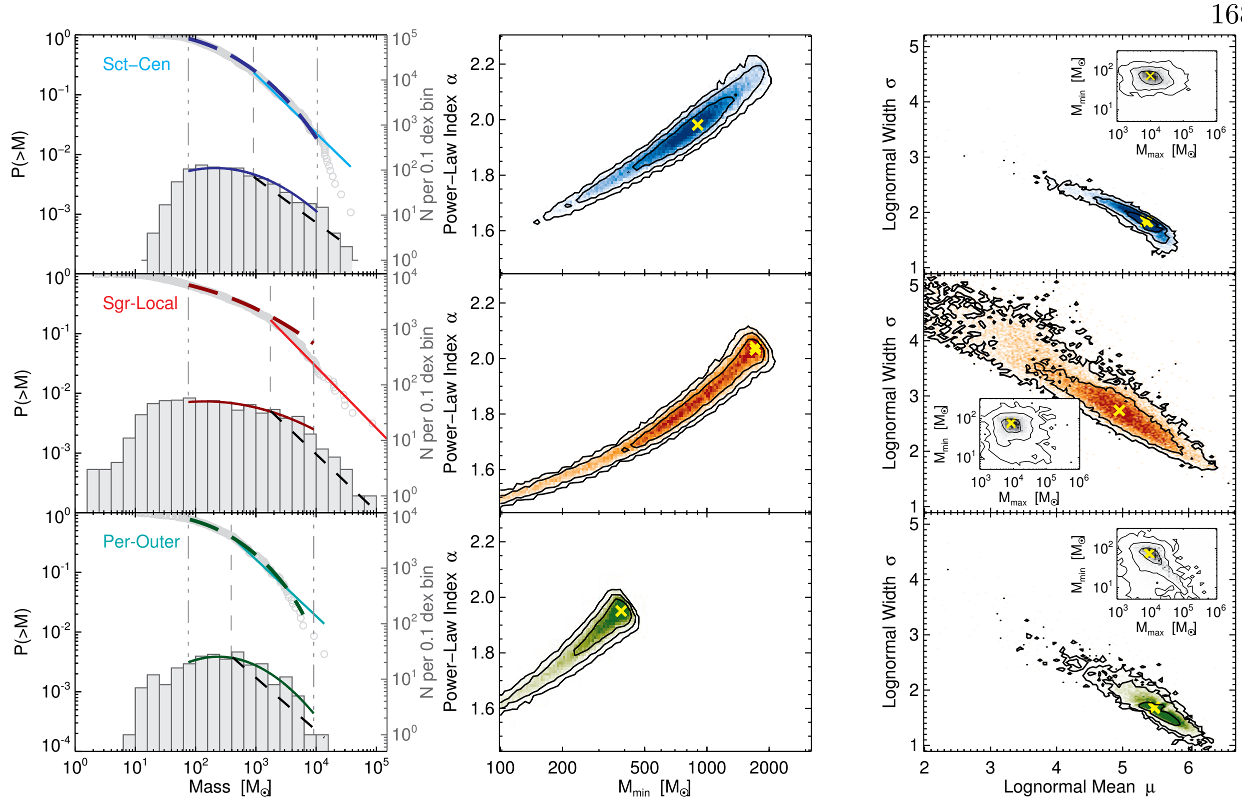

Figure 4.12 Mass functions of subsets in Galactocentric radius. Top row illustrates objects in the Scutum-Centarus arm ($R_{\rm gal} \leq 5.5$ kpc; lines in blue), the middle row objects in the Sagittarius and Local arms (5.5 kpc $\leq R_{\rm gal} \leq 8.5$ kpc; lines in red), and the bottom row objects in the Perseus and Outer arms (lines in green). Columns are as in Figure 4.11.

nicutt, 1998). We use Galactocentric radius as a proxy for Galactic environment to study the differences in the mass distribution from the Molecular Ring out to towards the truncation of the stellar disk ($R_{\rm gal} \approx 15$ kpc; Ruphy et al., 1996). It is somewhat difficult to disambiguate the effects of heliocentric distance from Galactocentric radius due to the Sun's location within the Galactic disk (**i.e.**, objects with small $R_{\rm gal}$ have larger $d_\odot$; see Figure 4.6 for the effects on the physical properties), and we analyze the systematic effects inherent therein. The histogram of source $R_{\rm gal}$ (Figure 4.10, **center left**) shows several concentrations of sources that provide useful subsets for investigation. The molecular ring / Scutum-Centarus Arm feature contains objects at $R_{\rm gal} \leq 5.5$ kpc (52%), sources at 5.5 kpc $\leq R_{\rm gal} \leq 8.5$ kpc are likely part of either the Sagittarius or Local Arms (35%), and anything beyond the solar circle ($R_{\rm gal} \geq 8.5$ kpc) is part of either the Perseus or Outer Arms (13%). This assignment is not one-to-one, as seen in the object-location face-on view of the

Milky Way from EB14 (their Figure 14) where some objects apparently identified with the Perseus arm on the far side of the tangent point near $30° \lesssim \ell \lesssim 50°$ lie within the middle $R_{\rm gal}$ group. For the purposes of this analysis, the particular arm to which an object belongs is subordinate to the distance from the Galactic center; arm names are simply used as descriptive identifiers.

The individual mass distributions and functional fits to each subset are shown in Figure 4.12; panel ordering and content is as in Figure 4.11. Objects in the Scutum-Centarus arm (top row) illustrate a somewhat-constrained power-law fit with $\hat{\alpha} = 1.97^{+0.08}_{-0.09}$ for $M \gtrsim 10^3\,M_\odot$, although the same high-mass steepening at $M \gtrsim 10^4\,M_\odot$ found for the complete sample is suggested in the CCDF (gray circles). The lognormal fit for this subset is characterized by $(\hat{\mu},\,\hat{\sigma}) = (4.9,\,1.7)$, with a tail extending towards lower μ, higher σ, as is a general feature of these fits. For this subset, neither functional form yields tight parameter constraints, suggesting the existence of substantial remaining inhomogeneities.

The middle row of Figure 4.12 illustrates the intermediate $R_{\rm gal}$ sources, mostly associated with the Sagittarius and Local arms. The wide spread in the mass distribution is a consequence of the large range of $d_\odot$ corresponding to this Galactocentric annulus. Although the maximum-likelihood power-law parameter values somewhat favor the high-mass end of the distribution ($\hat{\alpha} = 2.03^{+0.01}_{-0.27}$ for $M \gtrsim 1700^{+100}_{-1000}\,M_\odot$), the middle panel demonstrates a long tail of Monte Carlo trials to low-α. While this spread in the power-law fit parameters would seem to indicate a continuously curved mass distribution fit by a lognormal, similar to that in Figure 4.7 for the complete distribution, the middle-right panel of Figure 4.12 shows no particular localization of the lognormal parameters. The maximum-likelihood lognormal fit, $(\hat{\mu},\,\hat{\sigma}) = (4.95, 2.74)$, reasonably fits the wide plateau of the distribution, but the extremely heterogenous nature of this sample (non-localized joint likelihood contours) makes it less useful for drawing conclusions about dense molecular structures in the BGPS.

Objects in the Perseus and Outer arms beyond the Solar circle are depicted in the bottom row of Figure 4.12, where the small size of this subset ($N = 234$) compared to the others leads to the sharp truncation in the power-law parameters around $M_{\rm min} = 400\,M_\odot$, with a maximum-likelihood

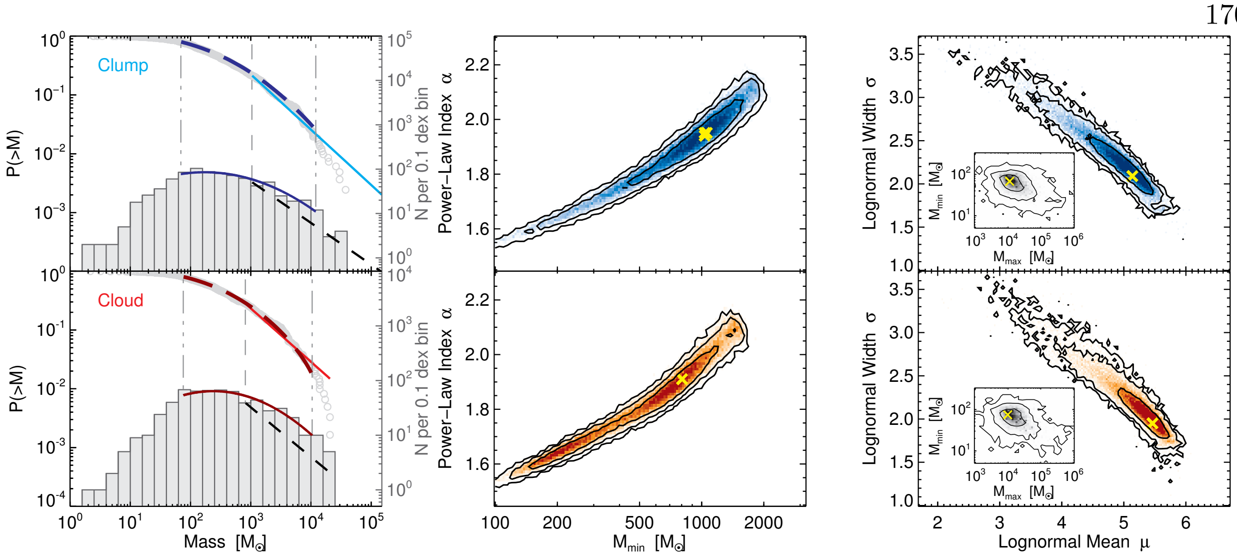

Figure 4.13 Mass functions of subsets in physical radius. Top row illustrates clump-scale objects ($R \leq 1.25$ pc; lines in blue), and the bottom row cloud-scale objects (lines in red). Columns are as in Figure 4.11.

$\hat{\alpha} = 1.94^{+0.04}_{-0.09}$. The lognormal fits (right panel) are constrained approximately as well as for the complete sample, but with a slightly larger mean. This subset is generally drawn from the outer Galaxy targeted fields in the BGPS, and likely is not completely representative of the outermost regions of the Galactic disk.

4.4.3.3 Clumps vs. Clouds

The derived physical radii and number densities of BGPS-detected dense molecular cloud structures divide the catalog into subsets more closely matched to the classification of the objects (Figure 4.10, **center right** and **far right**). There are relatively few objects falling in the "core" category ($R \leq 0.125$ pc or $n \geq 10^4$ cm^{-3}), and we lump these with the "clump" subset for this analysis. In terms of physical radius, "cores" all have small masses ($M \lesssim 40\,M_\odot$), but the collection of "cores", as defined by number density, have masses which range up to $10^4\,M_\odot$ (likely a consequence of uncertainties in source geometry; see Figure 4.5). We therefore divide sources into "core/clump" and "cloud" groups; the physical radius subsets are divided into core/clump (57%) and cloud (43%) sets, while cutting along number density yields the more-lopsided core/clump

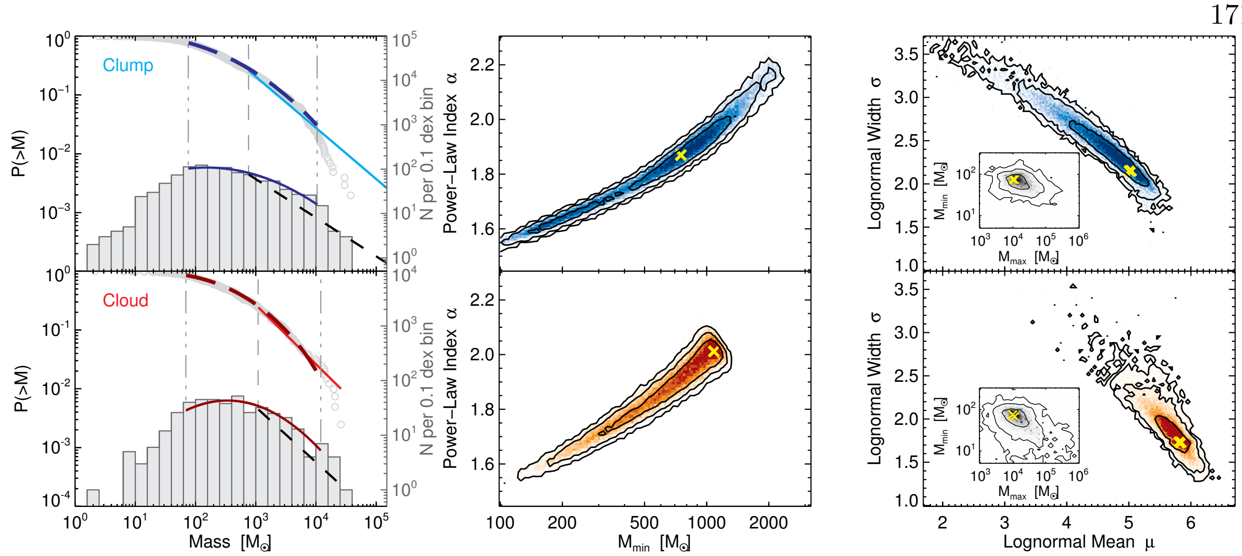

Figure 4.14 Mass functions of subsets in number density. Top row illustrates clump-scale objects ($n \geq 750$ cm^{-3}; lines in blue), and the bottom row cloud-scale objects (lines in red). Columns are as in Figure 4.11.

(72%) and cloud (28%) sets. The lack of a one-to-one mapping between these two physical tracers can be seen in Figure 4.5 (**middle**), where many of the large-spatial extent ($R \geq 1.25$ pc) objects are still somewhat dense ($n \geq 750$ cm^{-3}). There are likely observational biases present to skew this relationship, including the blending of substructure into larger-radius complexes when convolved with the telescope beam (see Merello et al., 2014).

Functional fits to the mass distributions are shown in Figures 4.13 (physical radius cuts), and 4.14 (number density cuts). Given the similarities between the mass distributions for the two methods of classifying dense molecular cloud structures, we discuss them together. The clump-scale objects are shown in the top rows of Figures 4.13 and 4.14. For both physical property criteria, the power-law fits stretch over wide range of parameter space with indices that span a range $1.6 \lesssim \alpha \lesssim 2.2$, although both have a maximum-likelihood value near $\hat{\alpha} = 1.9$. The lognormal fits also return similar values for the two clump criteria, with peaks at slightly lower mass and widths slightly larger than the lognormal fit to the complete data set.

The two cloud criteria also show great similarity in their functional fits (bottom row of Figures 4.13 and 4.14). The primary difference in the power-law fits comes from the smaller sample size for number density group ($N = 406$ for clouds with $n < 750$ cm^{-3} versus $N = 608$ for the clouds with $R > 1.25$ pc) leading to the truncation of the joint confidence intervals at large $M_{\rm min}$ in the bottom middle panel of Figure 4.14. The lognormal fits are very similar, both to each other and to the clump fits in the top rows of these figures. These sets of fits suggest that dividing the dense molecular cloud structures into categories based on physical radius or number density alone do not provide homogenous or distinct subsets for which meaningful mass functions may be fit that help constrain the theories of molecular cloud evolution.

4.4.3.4 Attempts to Find a Protocluster Mass Function

In addition to the strict cuts along the physical quantities described in Figure 4.10, additional subsets of the data were constructed in an attempt to define a "protocluster" mass function. The subsets are shown in Figure 4.15 along with their respective functional fits, and are discussed below.

The first two subsets are the blindly surveyed portion of the BGPS ($\ell \leq 90°$) and all objects with physical radius $R \geq 0.2$ pc. The outer Galaxy portions of the BGPS were targeted specifically based on known regions of star formation or concentrations of emission in the ^{12}CO(1-0) maps of Dame et al. (2001), meaning they are typically nearby clouds and suffer significant selection bias. Utilizing sources only from the inner Galaxy, where survey coverage was not (generally) predicated on prior knowledge of the structure of emission,[12] allows for a more unbiased sampling of sources (and therefore mass distribution) for fitting a mass function. The second subset removes all core-scale objects (which typically have $M < 100\,M_\odot$, see Figure 4.5) and those with non-finite θ_R (which have low flux-density and therefore likely low mass, see §4.4.1.1). Of course, the removal of low-mass objects will not appreciably affect the high-mass end of the mass distribution, and both of these subsets have power-law fit parameters $\hat{\alpha} \approx 1.9$, $M_{\rm min} \approx 900\,M_\odot$, consistent with the fit parameters for the complete set, and the loci of points in middle panels of Figure 4.15 bear a strong resemblance to that in Figure 4.7. The lognormal fits similarly reflect the entire sample, and do

[12] The "generally" arises because of the several regions along the inner Galactic plane where the BGPS latitude coverage was flared out from the nominal $|b| \leq 0\overset{\circ}{.}5$ to $|b| \leq 1\overset{\circ}{.}5$ in areas with known large vertical extent of star-formation activity (see Aguirre et al., 2011).

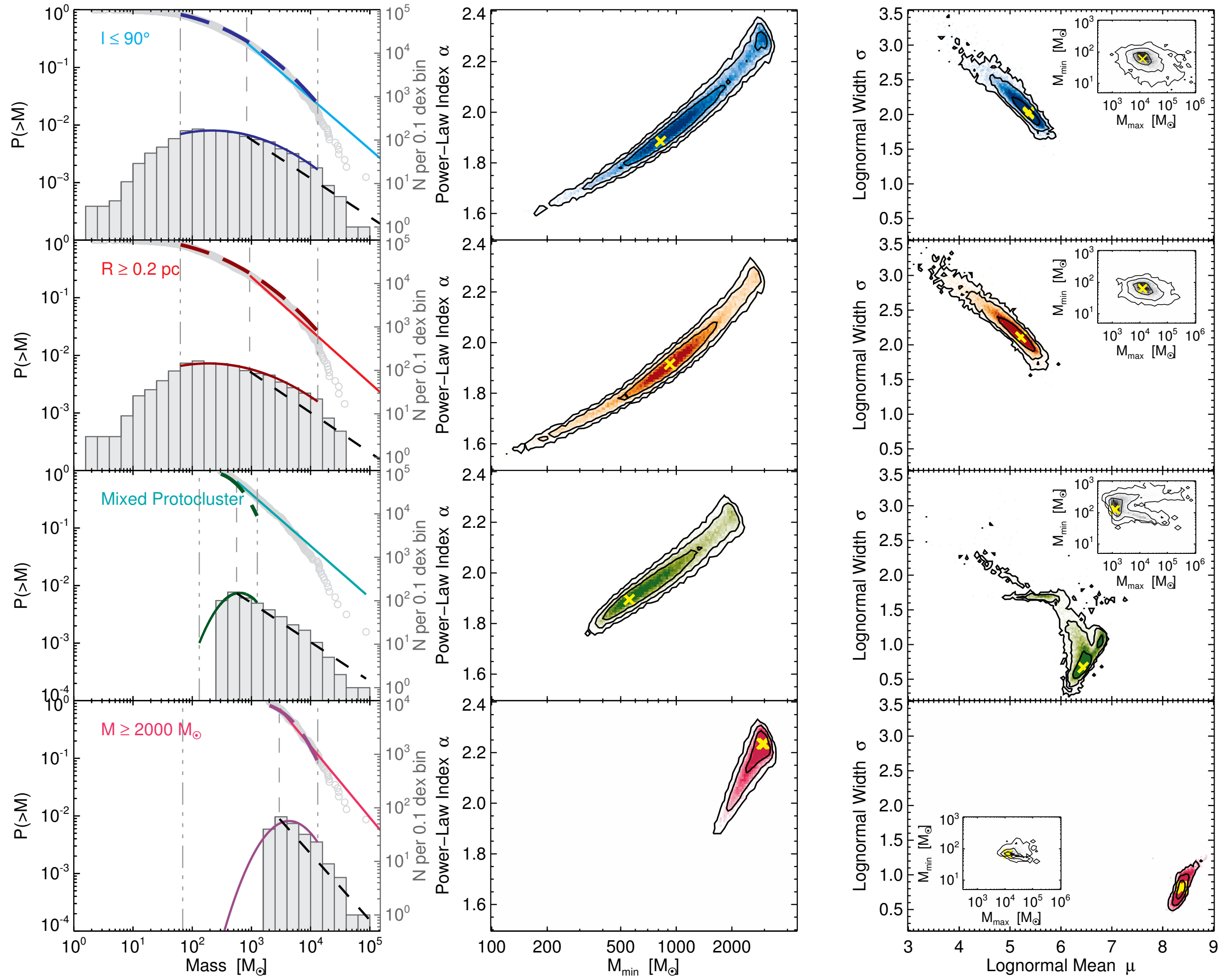

Figure 4.15 Mass functions of protocluster-type subsets. Top row only includes objects in the blind portion of the survey($\ell \leq 90°$; lines in blue), the second row objects of probable protocluster physical radius ($R \geq 0.2$ pc; lines in red), the third row objects with a mixed set of criteria (2 kpc $\leq d_\odot \leq$ 10 kpc and $M \geq 300\,M_\odot$; lines in green), and the bottom row the most massive objects ($M \geq 2000\,M_\odot$; lines in pink). Columns are as in Figure 4.11.

not add additional insight.

The other two subsets shown in Figure 4.15 begin to more strongly resemble power-law distributions. The third row, a mixed criterion in search of protoclusters, is the subset of sources with 2 kpc $\leq d_\odot \leq$ 10 kpc and $M \geq 300\,M_\odot$. This mix of conditions removes the most nearby objects (likely cores) and most distant objects (likely full clouds), and restricts consideration to a mass range consistent with star clusters (Kennicutt & Evans, 2012). The power-law fit joint confidence regions produce a narrow range of parameters with maximum-likelihood values at $\hat{\alpha} = 1.92^{+0.13}_{-0.06}$, $M_{\rm min} = 720^{+350}_{-250}\, M_\odot$. The power-law index is consistent with many of the fits, and extends to near the 90% mass completeness level. The lognormal fit displays an unusual set of contours, indicating that this functional form is not an appropriate description of the mass distribution. Finally, the bottom row of Figure 4.15 illustrates the fit to the massive end of the BGPS distribution ($M \geq 2000\, M_\odot$). The power-law fit parameters line up with the upper collection of points in the fit to the full sample, confirming the steep distribution ($\hat{\alpha} = 2.24^{+0.03}_{-0.07}$) for these sources. While the largest objects should conform to a power-law index $\alpha \approx 1.6 - 1.8$ (Blitz, 1993), we discuss in §4.5.3 some observational effects that might lead to the apparent steepening. The lognormal fit is tightly constrained, but since the minimum mass of this subset is substantially larger than the completeness levels, no real turnover in the data is expected.

4.4.4 Larson's Scaling Relationships

The physical properties of molecular clouds and their substructures appear to follow a series of scaling relationships known as Larson's Laws, following their introduction in a literature study of molecular clouds by Larson (1981). The most frequently used of these scaling relationships is the size-linewidth relationship, which holds that virial and/or turbulent motions across a molecular cloud structure increase with the square root of the size of that structure. This relationship is valid for single molecular line tracers (**e.g.**, Solomon et al., 1987; Heyer et al., 2009), but the linewidths of different molecular transitions in the same dense molecular cloud structure may vary by an order of magnitude. The heterogenous nature of the spectroscopic data gathered for BGPS distance estimation therefore prevents their use in deriving such a relationship.

The first two Larson relationships (the size-linewidth and mass-linewidth relationships) may be combined to form a mass-size relationship, sometimes called Larson's 3$^{\rm rd}$ Law.[13] While there has been some conjecture that this relationship is purely an observational artifact (**e.g.**, Ballesteros-

[13] This name is also given to the number density-size relationship similarly derived from the first two Larson relations.

Table 4.4. Power Law Fit Values for Larson's Mass-Radius Relationship

Data Set	$M = a\,R^b$ a	b	ρ^{a}
BGPS	317^{+20}_{-19}	1.87 (0.08)	-0.08
RD10	230^{+48}_{-39}	2.35 (0.09)	-0.90
BH14	525^{+253}_{-171}	2.00 (0.18)	-0.96
ALL	324^{+19}_{-18}	2.18 (0.04)	-0.59
Larson (1981)	460	1.90	⋯

[a]Parameter correlation coefficient $\rho = \sigma_{\rm ab}/\sigma_{\rm a}\sigma_{\rm b}$.

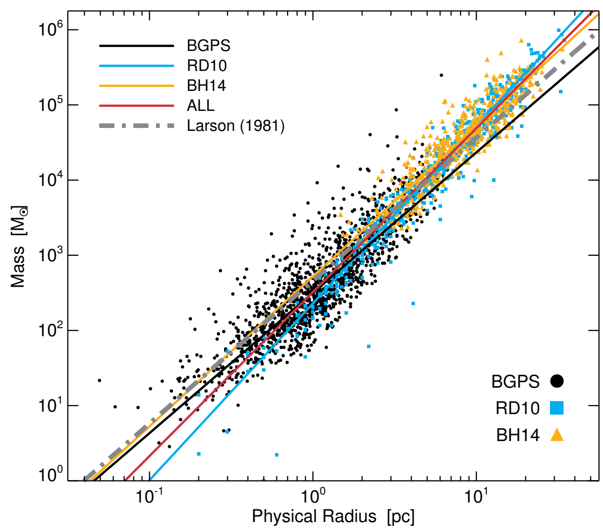

Figure 4.16 The mass-size relationship for molecular cloud structures. Black dots identify BGPS objects in Table 4.1, while cyan squares mark the values for GRS GMCs from Roman-Duval et al. (2010, labeled RD10) and orange triangles the GRS structures discussed in Battisti & Heyer (2014, labeled BH14). The black, cyan, and orange lines plot the respective power-law fits to the individual data sets, and the red line corresponds to the power-law fit for the combined data. The original result from Larson (1981) is shown as a thick gray dot-dashed line. Fit parameters for each line are listed in Table 4.4.

Paredes et al., 2012, and references therein), it is a generally-used relationship which has been shown to be valid in certain restricted settings (**e.g.**, Lombardi et al., 2010; Roman-Duval et al., 2010). Furthermore, Kritsuk et al. (2013) used simulations of isothermal supersonic turbulence to reproduce the observed relation.

Using the first two Larson (1981) relationships (the size-linewidth and mass-linewidth relationships), we derive

$$\frac{M}{M_\odot} = 460 \left(\frac{R}{\mathrm{pc}}\right)^{1.90} \quad , \tag{4.15}$$

where we use the physical radius $R = L/2$ rather than the overall length of the molecular feature for consistency with the remainder of our analysis. Equation (4.15) is based on measurements of GMCs using undersampled ^{12}CO(1-0) observations (see Heyer et al., 2009 for a comparison between the ^{12}CO data and the ^{13}CO(1-0) data from the GRS), and it is important to compare its prediction with millimeter-wave continuum observations of dense molecular cloud structures to assess its validity at different size scales.

Shown as black dots in Figure 4.16 are the mass and physical radius for the 1,427 BGPS Distance Catalog objects with finite θ_R. The black solid line is a linear fit in logarithmic space that is robust against outliers.[14] The result yields $M \propto R^{1.87}$, not far off from the Larson power-law $M \propto R^{1.90}$. We note that, systematically, $M \propto d_\odot^2$ and $R \propto d_\odot$, which leads to a similar $M \propto R^2$ relationship, but that theoretical underpinnings lead to the same conclusions (**e.g.**, Lombardi et al., 2010). To evaluate the correspondence of the relationship for the BGPS objects (**i.e.**, clump-scale structures) with larger ^{13}CO clouds from the GRS, we also plot in Figure 4.16 the GMCs studied by Roman-Duval et al. (2010, cyan squares) and the clouds containing BGPS sources from Battisti & Heyer (2014, orange triangles). In addition to the individual data sets, we also fit a power law through the entire collection of points (red) to span a further order of magnitude in each dimension than any one data set alone. The fit values and parameter correlations for each of the lines are listed in Table 4.4.

The BGPS and Battisti & Heyer (2014, BH14) fits both both have power-law indices consistent with the Larson value, but substantially shallower than the fit to the Roman-Duval et al. (2010, RD10) GRS data. The fact that the BH14 points generally lie above (at higher mass) than the BGPS fit line (black) indicates that there may be a systematic offset in the mass derivation between that from ^{13}CO and millimeter dust continuum.We note that although there are slight dif-

[14] Rather than a "least squares" fit, we utilized a "least absolute deviation" method to minimize the effects of outliers on the resulting fit (see Press et al., 2007, §15.7.3).

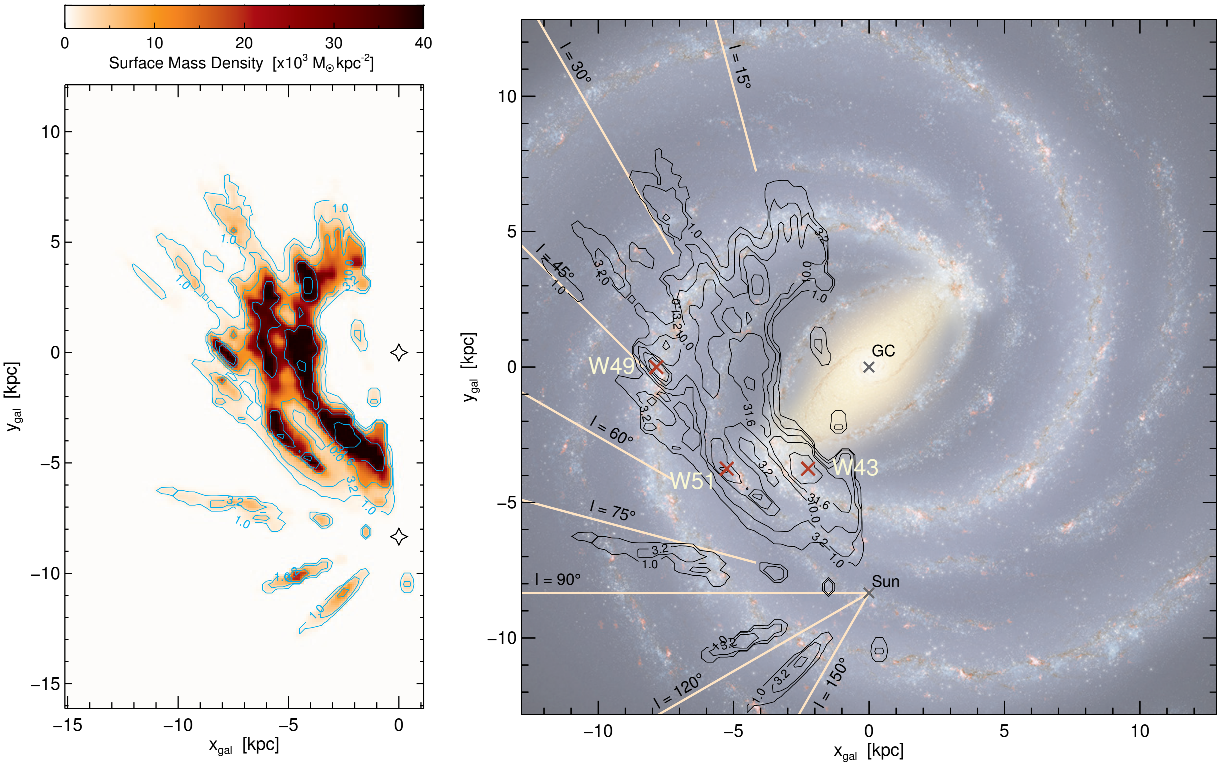

Figure 4.17 Dense gas mass surface density over the Galactic disk for Distance Catalog sources. **Left**: Color scale and contours indicate the dense gas surface mass density ($M_\odot$ kpc^{-2}) in each 0.25 kpc $\times$ 0.25 kpc pixel. The Sun and Galactic Center are identified by stars at $(x_{\rm gal}, y_{\rm gal}) = (0, -8.34)$ kpc and $(0, 0)$ kpc, respectively. **Right**: The same mass surface density contours plotted atop an artist's rendering of the Milky Way based on **Spitzer** data (R. Hurt: NASA/JPL-Caltech/SSC). Several of the mass concentrations correspond to prominent star-formation complexes originally discovered by Westerhout (1958, marked with red 'x's). As discussed in EB14, we have excluded from consideration of much of $R_{\rm gal} < 4$ kpc for kinematic reasons; sources appearing inside this circle have robust distance estimates through association with a trigonometric parallax measurement.

ferences in the fit parameters between the individual data sets, the power-law fit to the aggregate data (red line in Figure 4.16) holds remarkably well over more than five orders of magnitude in mass and two different tracers of molecular gas.

4.4.5 Distribution of Dense Gas in the Galactic Plane

4.4.5.1 Face-On View of the Milky Way

Ellsworth-Bowers et al. (2014, their Figure 14) presented a map of the locations of dense molecular cloud structures in the Galactic disk with respect to a rendering of Galactic structure based on **Spitzer** data. Since these objects span several orders of magnitude in mass (Figure 4.4), their locations alone do not necessarily indicate the mass distribution of dense molecular gas. To estimate the Galactic disk mass surface density of the objects in the BGPS Distance Catalog, we constructed a grid consisting of 0.25 kpc $\times$ 0.25 kpc bins into which masses computed for individual sources were placed. Since typical distance uncertainties are ~ 0.5 kpc, using 0.25 kpc bins captures the resolution of the data without smearing the image. Taking advantage of the information contained in source DPDFs, we produced a map (Figure 4.17, **left**) by randomly drawing 10^3 distances for each object, and adding the mass derived using Equation (4.3), divided by 10^3 to reflect the probabilistic contribution of that mass, to the appropriate grid cell. The use of the full DPDF leads to some of the smearing effects along the line of sight towards isolated regions in the outer Galaxy (**e.g.**, $\ell \approx 110°, 130°$). Furthermore, given that BGPS coverage at $\ell > 90°$ is neither blind nor contiguous, lack of mass in Figure 4.17 in the outer Galaxy should not be construed to **mean** a void in the Galactic plane.

The surface-density contours are plotted atop an image of the Galaxy produced using **Spitzer** near-infrared data (R. Hurt: NASA/JPL-Caltech/SSC) in the right panel of Figure 4.17. As with the corresponding map from EB14, there are many caveats in the interpretation of this image. First of all, there are virtually no BGPS sources with well-constrained distances at $R_{\rm gal} \lesssim 4$ kpc owing to non-circular motions induced by the Galactic bar,[15] except for objects associated with a trigonometric parallax measurement of maser transitions in regions of high-mass star formation (**cf.** Reid et al., 2014). Second, given that the typical uncertainty for well-constrained distance estimates from EB14 is ~ 0.5 kpc, the map and contours have a ~ 1 kpc effective resolution, significantly coarser than the underlying image, although commensurate with **Herschel** maps of nearby galaxies

[15] The DPDF formalism relies primarily upon kinematic distances whereby velocity measurements of dense molecular gas tracers are mapped onto Galactic position using a rotation curve assuming circular motion about the Galactic center. The Galactic bar induces strong radial streaming motions, rendering regions under its influence unusable for kinematic distance measurement.

(mean distance ≈ 10 Mpc; Kennicutt et al., 2011). Third, there is significant concentration of mass near $(x_{\rm gal}, y_{\rm gal}) = (-4, 3)$ kpc that appears to lie in a void in the underlying image. As discussed in EB14, this could represent either molecular gas at that location not identified in the **Spitzer** data (which is a "best guess" picture of the Galaxy), or sources in the W43 star-forming complex at the end of the Galactic bar erroneously placed at the far kinematic distance.

Comparing Figure 4.17 with the results of Roman-Duval et al. (2010) for the Galactocentric locations of GRS GMCs, there is a similar massive collection of gas at this location (their Figure 5). Since most of the BGPS sources in this region are associated with one or more H II region with robust KDA resolution from the H II Region Discovery Surveys (HRDS; Bania et al., 2010, 2012) and the distances to GRS clouds were primarily resolved using the same H I absorption techniques as the HRDS, the BGPS-GRS correlation implies either the existence of a large collection of mass in an apparently empty region of the **Spitzer** image or a fundamental limitation in the application of H I absorption techniques for distance resolution. Evidence in support of the former comes from both Russeil et al. (2011), who find interarm clumps for this region in the Hi-GAL $\ell = 30°$ SDP field, and a study of the M51 spiral galaxy that reveals a population of GMCs downstream of a major spiral arm (Egusa et al., 2011). Furthermore, the appearance of interarm gas could also be an effect of a spiral feature exhibiting a systematic noncircular velocity, biasing kinematic distances.

Identified in the right panel of Figure 4.17 are the massive-star forming regions W43, W51 and W49, which stand out in the mass diagram as high surface-density areas. Despite the rough tracing of some spiral structure, this map is not a definitive distribution of dense molecular gas in the Milky Way. First of all, objects used to produce this map represent only 20% of the full BGPS catalog (by number, although 33% by flux density). Secondly, given the coarse resolution of the map, any conclusions about detected spiral structure should be used cautiously as the map serves to show where the detected mass lies, and not a complete census of mass in the Galaxy. The expansion of robust distance determination to other dust-continuum surveys, however, will have high utility for tracing Galactic structure across the entire disk.

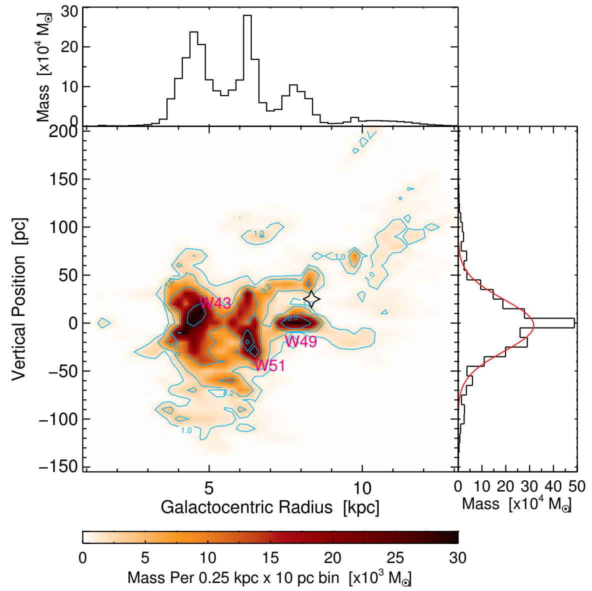

Figure 4.18 Vertical and radial mass distributions. The color-scale image in the center represents the azimuthal integration of mass, computed in analogous fashion to Figure 4.17, **left**. The absolute value of the color-scale is not physically meaningful, but represents the relative concentrations of mass. At top is the mass distribution versus $R_{\rm gal}$, marginalizing over vertical position, and at right is the distribution versus z, marginalizing over Galactocentric radius. The fit to the mass-weighted z histogram is consistent with the number-count histogram from Ellsworth-Bowers et al. (2014).

4.4.5.2 Vertical and Radial Distributions of Star-Forming Mass

As an alternative view of Galactic star-formation distribution, the azimuthally integrated vertical and radial distributions of the mass of star-forming clouds are shown in Figure 4.18. The Galactocentric radius and vertical position were computed using the $(\ell, b, d_\odot) \rightarrow (R_{\rm gal}, \phi, z)$ conversion matrix from Appendix C of EB13. As in Figure 4.17, masses and Galactocentric positions were derived from 10^3 independent samples from the DPDF for each object. The color image at the center represents the amount of mass in each 0.25 kpc (radial) $\times$ 10 pc (vertical) bin. Since this is an azimuthal integration, the color scale intensity of the image is meaningless other than to identify overdense regions; pixels at larger $R_{\rm gal}$ represent a significantly larger volume than those nearer the Galactic center. Identified in the image are the W43, W51, and W49 regions that stand out in the face-on image of the Galaxy. These three regions dominate the mass distribution, but there are still clear collections of mass in the 4-5 kpc range, and around 6 kpc. The warping of the disk to large $R_{\rm gal}$ is visible in the mass distribution, but we caution that BGPS observations in the outer Galaxy ($\ell > 90^\circ$) were targeted towards regions of known star formation, and do not represent the same blind sample as the inner Galaxy coverage.

To the right of the image is the marginalization over $R_{\rm gal}$, or the the vertical histogram of the mass distribution, shown as total mass per bin. A Gaussian fit to the this mass-weighted histogram ($\mu = -2.4 \pm 1.8$ pc, FWHM $= 63.7 \pm 3.6$ pc) is consistent with the distribution derived from the number-count histogram from EB14. This implies that the mass is roughly evenly distributed amongst catalog sources regardless of vertical position, and that there is no particular bias in the mass of dense molecular cloud structures near the Galactic midplane. We note that the BGPS survey latitude coverage does not fully probe the H_2 layer as measured by Bronfman et al. (1988), and that distance estimates for objects in the ATLASGAL or Hi-GAL surveys (both of which extend to $\pm 1^\circ$ as opposed to the nominal $\pm 0\overset{\circ}{.}5$ of the BGPS) would better answer the question of whether dense molecular gas settles near the Galactic midplane, but this is strong evidence that the scale height of the dense gas is smaller.

4.5 Discussion

4.5.1 What is a BGPS source?

Based on a subset of BGPS objects throughout the Galactic plane observed in the lowest inversion transitions of NH_3 around 24 GHz, Dunham et al. (2010, 2011b) began a discussion about the nature of BGPS sources. Using the larger sample of objects with well-constrained distance estimates from EB14, we continue this discussion.

Observationally, dense molecular cloud structures are typically defined in terms of cloud-, clump-, and core-scale objects (see §4.4.1.2). The BGPS is generally sensitive to clump-scale objects, but we investigate to what extent clouds and cores are present in the data. To illustrate the relative concentrations of non-clump objects at various heliocentric distances, Figure 4.19 shows the fraction of sources closer than a given distance with properties consistent with molecular cloud cores (thin back lines; $R \leq 0.125$ pc, $M \leq 27.5\,M_\odot$, and $n \geq 10^4$ cm^{-3}) based on physical radius

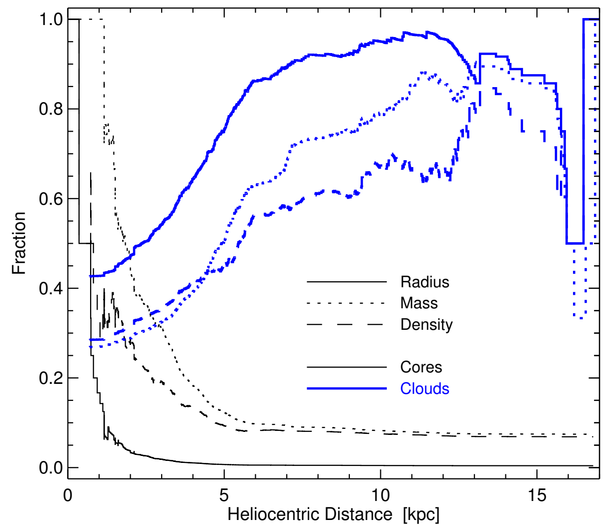

Figure 4.19 Fraction of source type as a function of heliocentric distance. Shown in thin black lines is the fraction of sources in the BGPS distance catalog nearer than the indicated heliocentric distance that are cores, and thick blue lines indicate the fraction of sources farther that are clouds. Line-styles indicate core and cloud divisions as determined by physical radius (solid lines), mass (dotted lines), and number density (dashed lines). Both ends of the distance range suffer from small number statistics. (This figure is styled after Dunham et al., 2011b, their Figure 21.)

(solid), mass (dotted), and number density (dashed). This figure is modeled after D11 (their Figure 21), and comparison between the two yields insight into the effects of using a larger sample. Many ($> 70\%$) of the objects at $d_\odot \leq 1$ kpc fall under the molecular cloud core definition, and conversely, the majority of core-scale objects are seen at $d_\odot \leq 2 - 3$ kpc. This is commensurate with the completeness and angular transfer functions illustrated in Figure 4.6.

Similarly, the fraction of sources located farther than a given distance that are characterized as clouds ($R \geq 1.25$ pc, $M \geq 750\,M_\odot$, and $n \leq 750$ cm^{-3}) are shown in Figure 4.19 as thick blue lines. Of particular note is the disparity in the distributions between different definitions for GMC-scale objects. Beyond 6 kpc, $\geq 90\%$ of objects in the Distance Catalog have large physical radius. As found by D11 for the NH_3-observed subset, the density curve never reaches 90% because there exist a substantial fraction of high-density sources in the distance range 6 kpc $\leq d_\odot \leq$ 13 kpc (see Figure 4.6, **bottom left**). Furthermore, the relative lack of tight correlation in the $M - n$ plot of Figure 4.5 (**right**) compared to the other plots in that figure indicates that the computed number densities of BGPS sources suffer from geometrical uncertainties in both the cataloging process and derivation of physical properties. The combination of cataloging routines with higher fidelity to the underlying structure and future high angular resolution studies of the Galactic plane (**e.g.**, CCAT) may alleviate this disparity. One principal deviation from the Dunham et al. results, however, is the mass fraction of cloud-scale objects as a function of distance. Whereas for the NH_3 subset the cloud fraction by mass closely tracked the radius fraction, in Figure 4.19 the cloud fraction by mass is significantly smaller.

With the enlarged set of physical properties presented here, we confirm the pervious conclusions of studies of BGPS objects that the majority of sources detected are clump-scale objects in the distance range 2 kpc $\lesssim d_\odot \lesssim$ 10 kpc, cores nearer by, and cloud-scale objects beyond. Uncertainties in the geometry of dense molecular cloud structures remains one of the most significant issues in terms of understanding the number density distribution and its role in the evolution of dense regions.

4.5.2 Larson's Laws & The Continuum of Structure

The continuum of structure within molecular cloud complexes is a fundamental property whose connection with the intimate processes of star formation requires further study. Much of the literature discusses objects as clouds, clumps, or cores, and assigns specific properties and physical significance to each (**cf.** McKee & Ostriker, 2007; Bergin & Tafalla, 2007; Kennicutt & Evans, 2012, and references therein). As seen in Figures 4.5 and 4.16, however, observed dense molecular cloud structures exhibit a continuum of properties across many orders of magnitude in scale; here we discuss the implications of this continuum.

As the relationship with the tightest correlation for the set of BGPS V2 objects studied here, we focus our attention on the mass-radius relationship. The combination of the size-linewidth and mass-linewidth relationships from Larson (1981) lead to $M \propto R^{1.90}$, and the empirical fits to the data in Figure 4.16 shown in Table 4.4 all reveal a relationship approximating $M \propto R^2$. Heyer et al. (2009) point out the basic definition

$$\frac{M}{M_\odot} = 100 \left(\frac{\Sigma}{100\, M_\odot \,\mathrm{pc}^{-2}} \right) \left(\frac{R}{\mathrm{pc}} \right)^2 \ , \tag{4.16}$$

and they measure surface densities for GMCs in the ^{13}CO(1-0) Galactic Ring Survey (GRS; Jackson et al., 2006) in the range $10\, M_\odot\, \mathrm{pc}^{-2} \lesssim \Sigma \lesssim 10^3\, M_\odot\, \mathrm{pc}^{-2}$, dependent on the size scale of the cloud. Furthermore, from excursion-set theory, Hopkins (2012) predicts $M \propto R^2$, again with a scale-dependent correction factor.

The confluence of theory and observation strongly indicate that this fractal quality of the dense interstellar medium[16] is perhaps more fundamental and a stronger driver of molecular cloud complex evolution than the individual classes of objects (**i.e.**, clouds, clumps, cores) working independently. Physical division are real, however, as the fractal nature of molecular clouds is driven by the the dominance of turbulence (Kritsuk et al., 2013); regimes dominated by other physical processes (**i.e.**, gravitational collapse) obey other scaling relationships.

Aside from questions of the evolution of the ISM, what are the implications of these scaling relations on observational data? As a basic reference point, we start from the Larson result of Equation (4.15) and convert the physical properties to observable quantities via Equations (4.3) and (4.5). The resulting relationship between observed flux density, deconvolved angular radius, and heliocentric distance is given by

$$\left(\frac{S_{1.1}}{\mathrm{Jy}} \right) = 3.34 \left(\frac{\theta_R}{\mathrm{arcmin}} \right)^{1.90} \left(\frac{d_\odot}{\mathrm{kpc}} \right)^{-0.10} \ . \tag{4.17}$$

This equation implies that for dense molecular cloud structures, the observed flux density is nearly independent of heliocentric distance, and depends only on the angular radius on the sky. This relationship is plotted in Figure 4.3 (**top**) as green lines for constant $d_\odot = 1$ kpc (solid), 5 kpc (dotted), and 20 kpc (dashed), and follows the data quite well. Therefore, in maps from (sub-)millimeter continuum surveys, the physical scales of detected objects are indistinguishable, meaning that a

[16] The interstellar medium, and molecular clouds in particular, display fractal dimensions (**e.g.**, Beech, 1987), wherein a self-similarity relationship exists over many orders of magnitude and no preferred scale exists between the limits of Galactic disk scale heights and runaway gravitational collapse of molecular cloud cores.

core looks like a clump looks like a cloud. It is only through robust distance estimation or application of the size-linewidth relationship that the scales of molecular cloud structures may be interpreted. This is an important point because the use of ancillary data is essentially the **only** means for deriving physically meaningful results from these surveys.

This last result may be applied to the analysis in EB14, where it was noted that the surface brightness (SB) distributions for the full BGPS catalog and various subsets (including the set of well-constrained distance estimates) were not so dissimilar as were the various flux density distributions. By combining the surface brightness equation from EB14 (their Equation 3) with Equation (4.17), we obtain

$$SB = 13.3 \text{ MJy sr}^{-1} \left(\frac{S_{1.1}}{\text{Jy}}\right)^{-0.05} \left(\frac{d_\odot}{\text{kpc}}\right)^{-0.105} , \tag{4.18}$$

which implies that the surface brightness of dense molecular cloud structures, regardless of size scale and heliocentric distance, should be nearly identical. Indeed, Solomon et al. (1987) comment that for the standard $\sigma_v \propto R^{0.5}$ size-linewidth relationship, virial equilibrium translates into constant average surface density, and hence constant average surface brightness. Given this, the observed surface brightness along any line of sight is then proportional to the number of structures, or total mass, along that line of sight. If one assumed an axisymmetric model and sampled a large fraction of the plane, the radial dependence (and normalization) of the molecular gas along the line of sight could be solved for.

In an orthogonal view on the continuum of structure in the dense interstellar medium, the Hi-GAL maps reveal a scale-independent combination of diffuse emission, filaments, and dense regions. This sequence is at least as fundamental as core-clump-cloud distinctions. The BGPS cataloging routine, however, it not designed to select on these criteria, so BGPS catalog objects do not span this axis, but are primarily compact regions; filamentary emission is frequently decomposed into multiple sources without information about the weblike structure visible in the **Herschel** maps.

4.5.3 Mass Distributions and Mass Functions

4.5.3.1 Aggregate Mass Distribution versus Subsets

The aggregate mass distribution from Figures 4.7 and 4.8 represents an inhomogenous set of dense molecular cloud structures across the Galactic plane. As such, it appears to be well-fit by a lognormal function over more than two orders of magnitude, and displays one or two power laws at the high-mass end. We made a series of astrophysically-motivated cuts in an attempt to isolate a more-homogenous population of sources that might enhibit a power-law function, the so-called clump mass function (**cf.** Donkov et al., 2012; Veltchev et al., 2013).

While the classifications of molecular cloud versus clump versus core are intuitively appealing, the mass distributions of BGPS sources divided into core/clump and cloud groups (Figures 4.13 and 4.14) return functional fits very similar to the aggregate set. Three possible explanations exist, likely working in concert: (1) the distinctions between object class is somewhat arbitrary, (2) uncertainties in the heliocentric distances mix objects across classification boundaries, and (3) the distinctions themselves are artificial in light of the continuum of structure in molecular cloud complexes. It is likely that a survey with improved angular resolution may be able to derive parameters as a function of properties. Furthermore, use of Hi-GAL will obviate the need to assume a dust temperature (or marginalize over the dust temperature, as done with the Monte Carlo trials), eliminating a source of uncertainty in derived properties.

Creating a subset primarily on cuts in heliocentric distance yielded the best power-law fits, characterized by nearly-symmetric joint confidence intervals in the middle panel of Figures 4.11-4.15. Both the "nearby" heliocentric distance subset and the "mixed protocluster" subset exclude the most distant objects, with the latter additionally excluding nearby, low-mass objects. While both are strongly power-law the "nearby" subset has a steeper power-law index ($\hat{\alpha} = 2.04 \pm 0.09$ vs. $1.92^{+0.13}_{-0.06}$), although the two are marginally consistent with each other.

The lack of a strong power-law fit for the "distant" sample or the two cloud-scale subsets (in R and n) may be partially caused by the apparent steepening in the mass distribution at high-mass (see Figure 4.15, **bottom**), an effect that is discussed below. The subsets that **do** exhibit power-law structure, however, indicate that the presence of a true underlying power law (if present) is not obscured and converted to a lognormal by sample inhomogeneity.

4.5.3.2 Steepness of the High-Mass Power-Law Fit

One of the curious aspects of the mass function fits is the propensity for the high-mass end to display a fairly steep power-law tail ($\alpha \approx 2.2 - 2.3$). While this power-law index is consistent with the high-mass end of the Salpeter (1955) IMF, objects detected by the BGPS in this mass range are generally cloud-scale (see Figure 4.5). Since Galactic GMCs tend to follow a power-law mass function with $\alpha \approx 1.6 - 1.8$ (Blitz, 1993), the present result is somewhat contradictory.

A steep power-law function (**i.e.**, $\alpha > 2$) implies that most of the total mass is in low-mass objects and that high-mass objects are rarer than more shallow functions (**i.e.**, there are fewer per mass bin; Equation 4.9). Could the apparent steepening be caused by "missing" high-mass sources in the present distance catalog? Removal of sources from a shallow power-law distribution does have the effect of imitating a steeper power law, given that a progressively larger fraction of sources are removed with increasing mass.

Observationally, the BGPS removes emission on large angular scales ($\gtrsim 300''$; G13). Additionally, the BGPS cataloging routine, based on a seeded-watershed algorithm, tends to break up large-scale emission into smaller chunks based on intermediate peaks in the flux-density distribution. What effect do these have on the derived mass distribution? Converting the mass-size scaling relationship of Equation (4.15) into observational quantities via Equation (4.5) yields

$$M = 43.8\, M_\odot \left(\frac{\theta_R}{\mathrm{arcmin}}\right)^{1.90} \left(\frac{d_\odot}{\mathrm{kpc}}\right)^{1.90} . \qquad (4.19)$$

Since the BGPS angular transfer function essentially limits the angular radius of objects to $\theta_R \lesssim 2\overset{\prime}{.}5$, large clouds will be filtered out or broken into substructures in the catalog, leading to an apparent lack of high-mass objects. Furthermore, the breaking up of large structures into smaller ones adds objects to the low-mass end of the mass distribution, steepening the power law. Therefore, accurate measurement of the high-mass power-law distribution of dense molecular cloud structures requires retaining some amount of large-scale structure in the survey images (**e.g.**, Hi-GAL).

4.5.3.3 Comparisons with the Literature

Placing the present mass function results in the context of other measurements for dense molecular cloud structures is very important. Whereas there is much discussion in the literature about dense molecular cores and their relationship with the stellar IMF, it is only the recent advent of large-scale Galactic plane surveys that have made available large, statistically robust samples of clump-scale objects. The only other study to date of a sizable number of clump-scale objects is Olmi et al. (2013), using the science demonstration fields from the **Herschel** Hi-GAL survey (Molinari et al., 2010b).

Comparison with the Olmi et al. (2013) results, however, is not completely direct, as the source-finding algorithm used by the Hi-GAL team is based on curvature from the second-derivatives of the image (Molinari et al., 2011) and typically finds structures on scales smaller than those identified in the BGPS catalog. Furthermore, Hi-GAL sources were identified as well-separated sources visible in each of the $\lambda = 70$ μm, 160 μm, and 250 μm bands; the beam size at 250 μm is $17''$, approximately half the width of the BGS $33''$ beam. These two features imply Hi-GAL will find more, smaller sources than the BGPS. Indeed, comparing source lists, in the $\ell = 30°$ field, Olmi et al. report identifying 1,950 sources, whereas there are only 492 BGPS V2 sources in the same longitude range. While Hi-GAL does extend twice as far in latitude as the BGPS, the bulk of the **Herschel**-measured emission lies within the BGPS coverage. For the $\ell = 59°$ field, the differences are much more astounding; the BGPS catalog contains 17 sources but Hi-GAL finds 3,402. A visual comparison of the data sets reveals that the Hi-GAL tile contains significant filamentary structure at modest flux density whereas the BGPS mosaic exhibits nearly uniform noise, indicating that atmospheric removal has obliterated much of this low-level signal. The vast majority of the Hi-GAL sources here lie along these filaments (see Molinari et al., 2010a, their Figure 3). Careful comparison of Hi-GAL and BGPS maps, however, has revealed that much of what was thought to be atmospheric 1/f noise corresponds to the cirrus-like filamentary structure that has been made choppy by the filtering.

With the differences between the two data sets in mind, Olmi et al. (2013) find power-law fits of $(\hat{\alpha},\, M_{\rm min}) = (2.15,\, 212\, M_\odot)$ for the $\ell = 30°$ field and $(\hat{\alpha},\, M_{\rm min}) = (2.20,\, 7.3\, M_\odot)$ for the $\ell = 59°$ field. Since the Hi-GAL cataloging routine is finding sources much smaller and that the power-law indices of these samples are steeper than the BGPS, it is likely that those authors are measuring something closer to a core mass function rather than a clump-scale mass function. In the $\ell = 30°$ field, the lognormal fit has a peak near $M_\mu = e^{4.58} = 100\, M_\odot$, somewhat smaller than the fit to the entire BGPS set (Table 4.2). The strong preference for cores in the $\ell = 59°$ Hi-GAL field is apparent in the lognormal peak near $M_\mu = e^{0.91} = 2.5\, M_\odot$. A better understanding of these discrepancies will come when both data sets are analyzed with the same cataloging method to remove systematic effects introduced by source identification.

4.5.4 Estimating the Dense Gas Mass Fraction

The Galactic mass distribution in Figure 4.17 represents the surface density of the dense molecular gas observed by the BGPS across the Galactic plane. Although the mass in that figure is diluted into 0.25 pc $\times$0.25 pc pixels, it still represents a large-scale measurement of the dense gas mass on scales approximating the spatial resolution of external Galaxies in the **Herschel** KINGFISH survey at a mean distance for the sample of 10 Mpc (Kennicutt et al., 2011). Placing the Milky Way in the context of fully visible galactic disks is an important step towards linking small-scale processes observable locally with the global pictures gleaned from external galaxies. In addition to studying the mass surface density of dense molecular cloud structures, we can ask what the dense gas mass fraction is.

Because of the scales involved in our estimate of the dense gas mass surface density, it is important to select a model or measurement of the total molecular gas content at similar scales. For this purpose, we utilized the azimuthally averaged H_2 distribution modeled by Wolfire et al. (2003) from the ^{12}CO data presented in Bronfman et al. (1988). While this model does not take

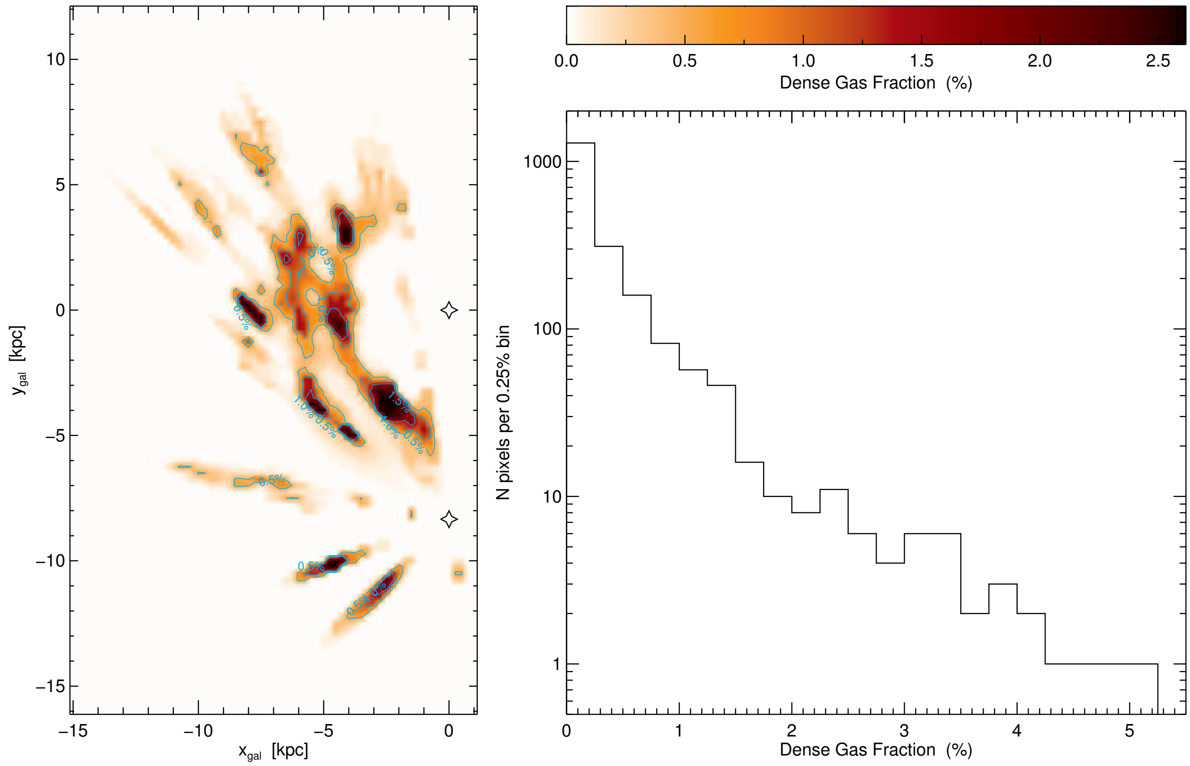

Figure 4.20 Dense gas mass fraction estimate as a function of Galactocentric position. **Left**: Face-on map computed by dividing the mass surface density of Figure 4.17 by the azimuthally averaged H_2 model of Wolfire et al. (2003). Contours indicate linear increments of 0.5%, and the stars mark the locations of the Galactic center $(x_{\rm gal}, y_{\rm gal}) = (0, 0)$ kpc, and the Sun $(x_{\rm gal}, y_{\rm gal}) = (0, -8.34)$ kpc. **Right**: Histogram of the non-zero pixels from the left panel. Most of the Galactic plane has $< 1\%$ measured dense gas mass fraction on these scales, but some pixels have $> 5\%$. The values shown are a lower limit – by as much as a factor of 5 – because this image includes only objects from the Distance Catalog.

into account azimuthal asymmetries such as spiral arms or the Galactic bar, it provides a baseline value to estimate the fraction of the surface density in the Galactic disk measured by the BGPS sources discussed here. The dense gas mass fraction map was computed by dividing the surface density in Figure 4.17 by the Wolfire et al. (2003) model, and is shown in Figure 4.20 (**left**) as percentages.

The right panel of Figure 4.20 shows the histogram of non-zero pixel values in the image. The vast majority of pixels have $< 1\%$ dense gas mass fraction, but some have $> 5\%$. Recall that the surface mass density in Figure 4.17 represents the mean of BGPS sources over 0.25 kpc $\times$ 0.25 kpc pixels, whereas the typical size of GMCs is ~ 10 pc. Therefore, the dense gas mass fraction shown in Figure 4.20 should be considered in the same light, as an average over a large area. Additionally, since the total mass shown comes from only 20% of the BGPS catalog (by number, but 33% by flux density), the values shown should be considered a lower limit on the dense gas mass fraction. By way of comparison, Battisti & Heyer (2014) compared the mass from BGPS sources with the mass of the parent GRS GMC, and found a mean value of $0.11^{+0.12}_{-0.06}$. Given the caveats about the nature of Figure 4.20 (**left**), the present measurement is consistent with the Battisti & Heyer value.

While the correspondence with the GRS value is encouraging, there are two issues that bear discussing. The first is "broad-brush" nature of the Wolfire et al. (2003) model of H_2 gas. Not only are there sizable uncertainties in both the disentangling of (optically thick) ^{12}CO emission into the near and far kinematic distances and the variable nature of the $X_{\rm CO}$ conversion factor as a function of environment (Bolatto et al., 2013), but enhancements in the surface density of H_2 in spiral arm features (coincident with the measured position of BGPS sources) would tend to decrease the measured dense gas mass fraction. Also, as previously stated, the severe dilution of GMC-scale objects into large pixels (based on the typical uncertainties in distance estimation; EB14) indicates that this measurement should be used for illustrative purposes only.

The second issue with interpreting this map is the finite resolution of current (sub-)millimeter continuum surveys and the nature of detected objects as a function of heliocentric distance. Beyond $d_\odot \approx 10$ kpc, the majority of detected objects are cloud-scale with lower mean density (see Figure 4.6). What is the threshold for objects to be included in the dense gas mass fraction? One illustration of this dilemma is the concentration of dense gas at $(x_{\rm gal}, y_{\rm gal}) \approx (-4, 3)$ kpc in Figure 4.20. While this may correspond to gas in the Sagittarius arm, is the BGPS detecting just the dense portion, or whole clouds ($d_\odot \approx 12$ kpc)? Despite these caveats, however, this type of measurement offers a key insight into the structure of the Milky Way within the context of galaxies in the local universe.

4.6 Summary

The true payoff of large-scale blind continuum surveys of the Galactic plane at (sub-)millimeter wavelengths is deriving the physical properties of regions hosting massive star formation. A detailed census of these dense molecular cloud structures can help constrain star-formation and galactic-evolution theories. In this work, we used the DPDF distance-estimation formalism of EB14 to compute the physical properties for the set of 1,802 BGPS V2 sources with well-constrained distance estimates. The significant aspects of this paper are:

(1) We utilize the full information available in the Bayesian DPDF framework introduced in EB13 and EB14 to compute physical property probability density functions. Although the DPDFs themselves are computed analytically on a linear distance scale, the application of the DPDF, in concert with probability density functions for source flux density with uncertainty and an estimate of the temperature distribution of sources, to the computation of physical properties requires Monte Carlo methods. We present a catalog of the masses, physical radii and number densities for the objects in the distance catalog of EB14, with uncertainties derived from the individual probability density functions.

(2) Unlike studies of isolated regions at a uniform distance, where flux-density completeness limits map directly onto mass completeness limits, large-scale blind dust-continuum surveys of the Galactic plane suffer from Malmquist bias and care must be taken to estimate the mass completeness levels for these heterogenous samples. We demonstrate a means for estimating the mass completeness function using the rms noise at the locations of sources in the survey. We estimate that the BGPS is 90% complete for $M \geq 400\,M_\odot$ and 50% complete for $M \geq 75\,M_\odot$.

(3) To understand the role of clump-scale features in molecular cloud complexes, we compute both power-law and lognormal fits to the mass distribution. Fits to the entire collection of sources yield a power-law index $\alpha = 1.9$ for objects with masses $10^3\,M_\odot \lesssim M \lesssim 10^5\,M_\odot$,

and a lognormal $(\mu, \sigma) = (5.3, 2.0)$, where the peak of the distribution is at $M_\mu = e^{5.3} = 200\,M_\odot$. We also fit both functional forms to a variety of astrophysically motivated subsets in an attempt to identify a collection of BGPS sources that represent a "clump mass function" or "protocluster mass function". For nearly all subsets we find power-law indices $1.85 \leq \alpha \leq 2.05$. The exception to this rule is the set of sources at $M \geq 2000\,M_\odot$, which has $\alpha \approx 2.2$; this steeper than expected power-law index may be caused by observation biases in the BGPS against large angular-extent sources. Lognormal fits to the subsets reveal that the more-inhomogenous collections of sources are better fit by this form. Collections of sources at similar distances show stronger power-law behavior. Furthermore, simulations of sources at a variety of heliocentric distances reveal that these inhomogenous samples produce an observed lognormal mass distribution with a power-law tail regardless of the underlying mass distributions.

(4) Combining the heliocentric distances from the DPDFs with the masses derived in this work, we compute the distribution of dense molecular gas in the Galactic plane. Both the face-on and azimuthally integrated $R_{\rm gal}$–z projections of the Galactic mass distributions reveal the prominent high-mass star-forming regions W43, W51, and W49 as significant concentrations of mass. The mass-weighted vertical mass distribution has a scale height ($\approx$ 30 pc) consistent with the number-count distribution from EB14, which is still about half that of ^{12}CO from Bronfman et al. (1988). Dividing the disk's mass surface density (face-on view) by a smooth model for Galactic H_2 distribution yields an estimate of the dense gas mass fraction, which peaks around 5% for the most massive complexes. This is consistent with the measured 11% value from the Battisti & Heyer (2014) study of GRS clouds, given that the BGPS mass is spread into pixels 250 pc on a side, whereas typical sizes of GRS clouds is $\lesssim$ 10 pc; our measurement is a lower limit due to the inclusion of only sources with well-constrained distances (21% of the BGPS catalog by number, 33% by flux density).

(5) The derived masses and physical radii of BGPS sources follow the extension of the mass-size relationship for GMCs first enumerated by Larson (1981), whereby $M \propto R^{\approx 2}$. There are several observational consequences of this relationship, principal of which is that the surface brightness of dense molecular cloud structures will be nearly constant independent of heliocentric distance. This means that the only limit on the detection of these objects in the Galactic plane are the finite resolution of extant telescopes at these wavelengths and the inescapable filtering out of large-scale emission in ground-based surveys owing to the subtraction of atmospheric signal from the raw data. This last point connects to the steepening of the power-law mass functions at large mass, as $M \propto \theta_R^{\approx 2}$ and filtering of large angular scale emission will eliminate some or most of the highest-mass objects in the Galactic plane.

The physical properties of BGPS-detected dense molecular cloud structures span the range from core to cloud, and illustrate the fractal nature of molecular cloud complexes, embodied in the Larson (1981) scaling relationships. While there are theoretical motivations for these designations, such as gravitational collapse for protoclusters (clumps) or individual stellar systems (cores), the systematic observational effects in current surveys makes it difficult to accurately categorize individual objects, especially since physical processes, such as gravitational binding, are environment-specific. What data such as those presented here can do, however, is begin to provide observational constraints on the evolution of molecular cloud complexes and the context for massive star formation.

Chapter 5

Concluding Thoughts

5.1 Summary

Can the forest be seen through the trees? In short, yes. (Sub-)Millimeter wavelength dust continuum surveys such as the Bolocam Galactic Plane Survey (BGPS) are able to peer through the optically thin dust emission and detect dense molecular cloud structures across the breadth of the Galactic disk. The fundamental fractal nature of molecular clouds (embodied in the mass-size scaling relationship), however, implies that objects of all size scales have approximately the same surface brightness. The "particular bugbear of astronomy" (Binney & Merrifield, 1998, p. 535) is thusly multiplied manyfold in that dense molecular cloud structures at various distances all **look** identical in the continuum images.

The problem, then, of distance estimation becomes absolutely critical for the deriving useful physics from the continuum images and placing observational constraints on the brief and poorly understood stage of massive star formation between the coalescence of giant molecular clouds and the emergence of protostars. This thesis developed and applied a novel Bayesian distance estimation method to the particular case of the BGPS. Both the method and accompanying software are completely general and may be applied to any similar survey of the Galactic plane. Following is a brief summary of each of the thesis chapters followed by some overall concluding remarks.

Chapter 2 introduced the distance probability density function (DPDF) framework for distance estimation, whereby a Bayesian posterior DPDF is computed as the product of a kinematic distance likelihood function and a series of prior DPDFs based on ancillary data. One of the chief difficulties in using kinematic distances in the inner Galaxy ($|\ell| \leq 90°$) is an ambiguity (KDA) between two distances along the line of sight that correspond to a single projected line-of-sight velocity. The purpose of prior DPDFs, then, is to resolve the KDA and place objects at a single distance. This chapter introduced two such priors. The first is based on the general distribution of H_2 throughout the Galactic plane, and constrains high-latitude ($|b| \gtrsim 0\overset{\circ}{.}4$) to be at the near kinematic distance, as these lines of sight quickly exit the layer of dense molecular gas near the Galactic midplane. The second utilizes the fact that the dust whose emission is detected at $\lambda = 1.1$ mm absorbs the ubiquitous mid-infrared emission from polycyclic aromatic hydrocarbons (PAHs) near $\lambda = 8$ μm. We developed a method for morphologically matching emission in the BGPS images with eight-micron absorption features (EMAFs) in the **Spitzer**/GLIMPSE mosaics.

Applying these two priors to the set of BGPS version 1 (V1) sources with measured velocities at the time from molecular transition line observations (principally HCO^+(3-2) and NH_3(1,1)), well-constrained distances (distance uncertainty $\lesssim 2$ kpc) were measured for 618 objects. Placement of these sources in a face-on view of the Milky Way revealed a concentration in the Molecular Ring / Scutum-Centarus Arm ($R_{\rm gal} \approx 4 - 5$ kpc) and hints of spiral structure associated with the Sagittarius Arm. Additionally, it was shown that EMAFs are a generalization of infrared dark clouds (IRDCs), objects associated with massive star formation. Previous studies have used IRDCs to assign dust-continuum sources to the near kinematic distance, but this chapter demonstrated that EMAFs may be seen at the far kinematic distance at lower contrast than their nearer cousins.

Chapter 3 expanded on the DPDF formalism by introducing a validated means for extracting a dense-gas spectrum from observations of ^{13}CO(1-0), which typically traces more diffuse gas, to derive kinematic distance likelihoods. A new prior DPDF was also presented that associates dust-continuum sources with robust distance estimates for star-forming complexes in the literature. This new prior was demonstrated using trigonometric parallax measurements for masers in high-mass star forming regions and robust KDA resolutions for H II regions associated with star formation. Furthermore, with the release of the re-reduction of the BGPS data (version 2), the entire DPDF framework was applied to the new data to produce a distance catalog containing 1,802 well-constrained sources (21% of the V2 catalog).

Analysis of the resulting distance catalog revealed several interesting aspects of the forest through the trees. The first is that dense molecular cloud structures apparently have a scale-height (≈ 30 pc) half that of the Galactic molecular layer as measured via ^{12}CO (≈ 60 pc; Bronfman et al., 1988). Second, portions of four spiral arms (Scutum-Centarus, Sagittarius, Perseus, and Outer) may be traced with BGPS objects, although the distance uncertainties (~ 0.5 kpc) inherent in the DPDFs somewhat smear out the spiral structures themselves.

Chapter 4 presented the astrophysical payoff of this work, whereby the BGPS distance catalog was used to compute physical properties for this collection of dense molecular cloud structures. The full information contained in the Bayesian DPDFs were used to create physical property probability density functions though a large number of Monte Carlo realizations. To connect the observed distribution of masses with theory, functional fits were performed for both power-law and lognormal forms through another series of Monte Carlo trials to identify the maximum-likelihood values of the fit parameters. These trials consisted of fitting functional forms to independent realizations of the mass distribution, wherein each object was represented by a mass drawn randomly from the previously computed probability density function. The maximum-likelihood power-law index for the high-mass end of the entire sample was $\hat{\alpha} = 1.94^{+0.16}_{-0.12}$, and fits to various astrophysically motivated subsets of the data yield power-law indices in the range $1.85 \leq \hat{\alpha} \leq 2.05$, intermediate between the slope for giant molecular clouds and the Salpeter (1955) stellar initial mass function.

The combination of mass and heliocentric distance allowed the creation of a mass-weighted face-on view of the Milky Way, wherein several massive star-forming regions (W43, W51, and W49) easily stand out. While the resolution of this map is too low to precisely trace spiral structure, it is commensurate with the recent **Herschel** KINGFISH survey of external galaxies (Kennicutt et al., 2011), which will allow the Milky Way to be placed in the direct context of other spiral galaxies. The mass-weighted vertical distribution of dense molecular gas is consistent with the number-count distribution from Chapter 3, suggesting a uniform distribution of mass amongst objects at different vertical heights above the midplane. Finally, the derived masses and physical radii demonstrate a continuation of the Larson (1981) scaling relationship, showing the fractal nature of the dense interstellar medium.

This thesis presents the first large-scale probabilistic methodology for measuring distances to and physical properties of dense molecular cloud structures in the Milky Way. The DPDF formalism can be thought of as "throwing everything but the kitchen sink" at the problem of distance determination because it allows for the application of an arbitrary number of prior DPDFs for any given object. Additionally, it may be extensible to computing distance probabilities as a function of position on the sky rather than rely on extracted catalog sources. The flexible nature of the accompanying software makes it directly applicable to current Galactic plane surveys (**e.g.**, ATLASGAL, Hi-GAL) and future, higher-resolution surveys from **e.g.**, CCAT.

The large collection of objects from across the Galactic plane represents the first large-scale analysis of clump-scale objects in a variety of Galactic environments. The mass function fits and physical property analysis presented in this thesis are the first steps on the road to seeing the molecular content of the Milky Way as a forest rather than individual nearby trees.

5.2 Future Work

The work presented in this thesis is but the tip of the iceberg in terms of measuring the structure of dense molecular gas in the Milky Way, both on Galactic- and cloud-based levels. Briefly described below are the future directions this work may be taken to capitalize on the foundation provided herein.

5.2.1 Additional Prior DPDFs

There is a long list of possible additional prior DPDFs that could be implemented in the service of robust distance estimation, but many of these will require much careful tuning and characterization before being production-ready. Each method is **very** briefly described, and references provided for further investigation.

Near-Infrared Extinction Molecular cloud clumps absorb starlight at short wavelengths. Because near-infrared light experiences less extinction from diffuse interstellar dust than optical or ultraviolet, it is able to probe larger distances through the Galactic plane. Foster et al. (2012) showed that BGPS objects may be seen by their effects on near-IR star count catalogs because of their large A_V, and Marshall et al. (2009) used similar methods to determine near-infrared extinction (NIREX) distances to over 1200 IRDCs. The underlying principle behind NIREX distances is the relatively narrow range of intrinsic near-infrared stellar colors (**i.e.**, $\Delta(J-K) \lesssim 1.5$ mag; Lombardi & Alves, 2001); a molecular cloud clump is therefore detectable through a bimodal histogram of the $J-K$ colors of stars within its boundaries.

There are two possible approaches to create a prior DPDF with this method. The first involves generating histograms of the $J-K$ colors of stars in catalogs of near-infrared surveys within the boundaries of a BGPS source and comparing the number of blue stars against the Besançon Galactic star count model (Robin et al., 2003) as a function of distance. The second involves creating smoothed images of each near-infrared band around each BGPS source for comparison to synthetic images for each band based on the cloud A_λ and the Besançon Galactic star count model in a means analogous to the EMAF method from Chapter 2.

21-cm Absorption by cold H I in Molecular Cloud Clumps Small amounts of neutral atomic hydrogen, maintained by cosmic ray dissociative ionization of H_2, exist in the cold central regions of molecular clouds. These atoms absorb background 21-cm emission that would otherwise pass unhindered though GMCs. This mechanism produces a narrow spectral shape owing to both the small turbulent velocities within molecular cloud clumps and the cold ambient temperatures. Use of H I self-absorption (HISA) and H I emission / absorption (HIE/A) techniques on a large scale for kinematic distance ambiguity resolution of star-forming regions was introduced by Roman-Duval et al. (2009) and Anderson & Bania (2009).

HISA occurs when cold H I on the near side of the tangent point absorbs diffuse H I emission from gas at the same $v_{\rm LSR}$ on the far side of the Galaxy. If an absorption feature is detected at the $v_{\rm LSR}$ of a dense molecular cloud structure, it must be at the near kinematic distance provided there is no 21-cm continuum emission at that location. To isolate HISA features associated with dense molecular cloud structures, an On-Off technique analogous to that presented in Chapter 3 for ^{13}CO is required. Likely spatial On-Off HISA maps (integrated over $v_{\rm LSR}$) will also need to be created to compare the morphology of the absorption with that of (sub-)millimeter emission following the reasoning in Chapter 2 for EMAFs.

The presence of 21-cm continuum emission requires the use of the HIE/A technique. In this case, the morphology of the dust continuum source is compared with the 21-cm continuum images to determine if the millimeter-wave object is correlated with the 21-cm continuum object. The continuum emission is checked for absorption features at all $v_{\rm LSR}$; if absorption features occur only for $v_{\rm LSR}$ less than that of the dense molecular cloud structure, the near distance is assigned, and if absorption features occur for $v_{\rm LSR}$ all the way up to the tangent velocity, then the far distance is assigned.

H I Column Densities In the outer Galaxy (Quadrants II and III), the distance to an object with a measured $v_{\rm LSR}$ may be estimated by comparing the observed H I column density (integrated from 21-cm observations out to $v_{\rm LSR}$) against a model of $N_{\rm HI}$ as a function of heliocentric distance (Foster & Routledge, 2003). This method is independent of the Galactic kinematic model chosen, since the $v_{\rm LSR}$ is used only to place bounds on the integration of H I 21-cm emission along the line-of-sight for comparison with the model. Being independent of kinematics, this method can place limits on the distances to objects in the Galactic longitude cardinal directions ($\ell = 90°, 180°$) where kinematic distances are unreliable because peculiar motions are on order the variation of the rotation curve on the scale of kiloparsecs. This method can only work where $v_{\rm LSR}$ maps 1:1 onto heliocentric distance (**i.e.**, it cannot be used to resolve the kinematic distance ambiguity for objects within the Solar Circle).

Size-Linewidth Relationship This is a prior DPDF based on Larson's Laws, which relate density, size, linewidth, and virial ratio for molecular clouds (Larson, 1981; Solomon et al., 1987; Heyer et al., 2009). Preliminary investigations of possible implementations use GRS ^{13}CO data to derive linewidth data cubes in the position-velocity vicinity of dense molecular structures.

Distances are derived by comparing sizes and linewidths of volumes of the ^{13}CO cubes (Rosolowsky et al., 2008b).

Spanning-Tree or Nearest-Neighbor Broadcasting This somewhat-nebulous idea involves applying the kinematic distance ambiguity resolution of a given dense molecular cloud structure to its neighbors in (ℓ, b,$v_{\rm LSR}$) space assuming that associations of dense molecular structures in (ℓ, b,$v_{\rm LSR}$) space correspond to associations in ($x_{\rm gal}, y_{\rm gal}, z_{\rm gal}$) space. Successful development of this prior could allow DPDFs to be broadcast to any dust-continuum-detected objects near those with kinematic velocities, increasing the well-constrained sample size severalfold.

5.2.2 Application of the DPDF Formalism to e.g., Hi-GAL and CCAT

Hi-GAL provides an unprecedented census of sites of Galactic star formation, mapping the entire Galactic plane with $|b| \leq 1°$ in five broad bands (70, 160, 250, 350, and 500 μm; Molinari et al., 2010b), making it the most comprehensive Galactic far-infrared + submillimeter survey in existence and is not likely to be superseded anytime in the foreseeable future. Not only is the spatial coverage uniquely comprehensive, but the wavebands encompass thermal dust emission spectral energy density (SED) peaks from 6 K to 41 K, bracketing the range of most Galactic star-forming regions (Olmi et al., 2013) and enabling robust temperatures and luminosities to be measured.

In order to identify discrete molecular cloud clumps in the presence of substantial diffuse emission (clumps are embedded within molecular clouds), the **Herschel**/SPIRE bolometer timestream data must be high-pass filtered before source finding. Adaptation of the Bolocat source-finding algorithm to identify sources for Hi-GAL will require tuning for higher S/N and lower 1/f noise than in the BGPS survey mosaics. Sources thusly extracted may be used with the DPDF formalism outlined in this thesis to produce distance and physical property estimates for dense molecular cloud structures across the entire Galactic plane.

Furthermore, the completion of CCAT[1] will allow significant leaps forward in the study

[1] This 25-m diameter telescope will be constructed at 5600 m altitude on Cerro Chajnantor in the Andes mountains

of the dense interstellar medium. The angular resolution in the prime 350-μm band will be $3\farcs5$, enabling a clear view of structures across a wide range of physical size. Data from CCAT will be able to clear up the cloud/clump/core confusion present in the current Galactic plane surveys, and the high sensitivity afforded by the large dish will be able to detect solar-mass objects across much of the Galactic plane.

of northern Chile. This exceptionally dry site will allow for observations in the 350-μm atmospheric window.

Bibliography

Aguirre, J. E., Ginsburg, A. G., Dunham, M. K., et al. 2011, ApJS, 192, 4, (A11)

Alves, J., Lombardi, M., & Lada, C. J. 2007, A&A, 462, L17

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

Ballesteros-Paredes, J., D'Alessio, P., & Hartmann, L. 2012, MNRAS, 427, 2562

Bally, J., Anderson, L. D., Battersby, C., et al. 2010a, A&A, 518, L90

Bally, J., Aguirre, J., Battersby, C., et al. 2010b, ApJ, 721, 137

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, ApJ, 718, L106

Barkana, R., & Loeb, A. 2001, Phys. Rep., 349, 125

Bartkiewicz, A., Brunthaler, A., Szymczak, M., van Langevelde, H. J., & Reid, M. J. 2008, A&A, 490, 787

Bastian, N., Covey, K. R., & Meyer, M. R. 2010, ARA&A, 48, 339

Battersby, C., Bally, J., Jackson, J. M., et al. 2010, ApJ, 721, 222

Battersby, C., Bally, J., Ginsburg, A., et al. 2011, A&A, 535, A128

Battisti, A. J., & Heyer, M. H. 2014, ApJ, 780, 173, (BH14)

Bayes, M., & Price, M. 1763, Philosophical Transactions, 53, 370

Beech, M. 1987, Ap&SS, 133, 193

Benjamin, R. A., Churchwell, E., Babler, B. L., et al. 2003, PASP, 115, 953

—. 2005, ApJ, 630, L149

Bergin, E. A., & Tafalla, M. 2007, ARA&A, 45, 339

Binney, J., & Merrifield, M. 1998, Galactic Astronomy (Princeton University Press)

Blaauw, A., Gum, C. S., Pawsey, J. L., & Westerhout, G. 1960, MNRAS, 121, 123

Blitz, L. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine, 125–161

Bolatto, A. D., Wolfire, M., & Leroy, A. K. 2013, ARA&A, 51, 207

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. 2009, ApJ, 693, 424

—. 2011, AN, 332, 461

Burkert, A., & Hartmann, L. 2004, ApJ, 616, 288

Burton, M. G., Braiding, C., Glueck, C., et al. 2013, PASA, 30, 44

Butler, M. J., & Tan, J. C. 2009, ApJ, 696, 484

Chabrier, G. 2003, PASP, 115, 763

Churchwell, E., Babler, B. L., Meade, M. R., et al. 2009, PASP, 121, 213

Clauset, A., Shalizi, C. R., & Newman, M. E. J. 2009, SIAM Review, 51, 661

Clemens, D. P. 1985, ApJ, 295, 422

Clemens, D. P., Sanders, D. B., & Scoville, N. Z. 1988, ApJ, 327, 139

Cohen, R. S., Cong, H., Dame, T. M., & Thaddeus, P. 1980, ApJ, 239, L53

Contreras, Y., Schuller, F., Urquhart, J. S., et al. 2013, A&A, 549, A45

Dame, T. M. 2011, ArXiv e-prints, arXiv:1101.1499

Dame, T. M., Hartmann, D., & Thaddeus, P. 2001, ApJ, 547, 792

Dame, T. M., & Thaddeus, P. 2008, ApJ, 683, L143

—. 2011, ApJ, 734, L24

Dame, T. M., Ungerechts, H., Cohen, R. S., et al. 1987, ApJ, 322, 706

Dempsey, J. T., Thomas, H. S., & Currie, M. J. 2013, ApJS, 209, 8

Dobbs, C. L., & Burkert, A. 2012, MNRAS, 421, 2940

Dobbs, C. L., Krumholz, M. R., Ballesteros-Paredes, J., et al. 2013, ArXiv e-prints, arXiv:1312.3223

Donkov, S., Veltchev, T. V., & Klessen, R. S. 2012, MNRAS, 423, 889

Draine, B. T. 2003, ARA&A, 41, 241

Draine, B. T., & Li, A. 2007, ApJ, 657, 810

Dunham, M. K., Robitaille, T. P., Evans, II, N. J., et al. 2011a, ApJ, 731, 90

Dunham, M. K., Rosolowsky, E., Evans, II, N. J., Cyganowski, C., & Urquhart, J. S. 2011b, ApJ, 741, 110, (D11)

Dunham, M. K., Rosolowsky, E., Evans, II, N. J., 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., Rosolowsky, E. W., Glenn, J., et al. 2014, ApJ, submitted, (EB14)

Ellsworth-Bowers, T. P., Glenn, J., Rosolowsky, E., et al. 2013, ApJ, 770, 39, (EB13)

Elmegreen, B. G. 1985, in Birth and Infancy of Stars, ed. R. Lucas, A. Omont, & R. Stora, 257–277

Enoch, M. L., Glenn, J., Evans, II, N. J., et al. 2007, ApJ, 666, 982

Enoch, M. L., Young, K. E., Glenn, J., et al. 2006, ApJ, 638, 293

Evans, II, N. J. 1999, ARA&A, 37, 311

Ewen, H. I., & Purcell, E. M. 1951, Nature, 168, 356

Foster, J. B., Stead, J. J., Benjamin, R. A., Hoare, M. G., & Jackson, J. M. 2012, ApJ, 751, 157

Foster, T., & Routledge, D. 2003, ApJ, 598, 1005

Fux, R. 1999, A&A, 345, 787

Georgelin, Y. M., & Georgelin, Y. P. 1976, A&A, 49, 57

Giannetti, A., Brand, J., Sánchez-Monge, Á., et al. 2013, A&A, 556, A16

Gibson, S. J., Taylor, A. R., Higgs, L. A., Brunt, C. M., & Dewdney, P. E. 2005, ApJ, 626, 195

Ginsburg, A., Glenn, J., Rosolowsky, E., et al. 2013, ApJS, 208, 14, (G13)

Glenn, J., Ade, P. A. R., Amarie, M., et al. 2003, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 4855, Millimeter and Submillimeter Detectors for Astronomy, ed. T. G. Phillips & J. Zmuidzinas, 30–40

Goldsmith, P. F., Li, D., & Krčo, M. 2007, ApJ, 654, 273

Gómez, L., Wyrowski, F., Schuller, F., Menten, K. M., & Ballesteros-Paredes, J. 2014, A&A, 561, A148

Herschel, W. 1800, Royal Society of London Philosophical Transactions Series I, 90, 284

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

Hildebrand, R. H. 1983, QJRAS, 24, 267

Honma, M., Nagayama, T., Ando, K., et al. 2012, PASJ, 64, 136

Hopkins, P. F. 2012, MNRAS, 423, 2016

—. 2013, MNRAS, 433, 170

Humphreys, R. M., & Larsen, J. A. 1995, AJ, 110, 2183

Indebetouw, R., Mathis, J. S., Babler, B. L., et al. 2005, ApJ, 619, 931

Jackson, J. M., Finn, S. C., Rathborne, J. M., Chambers, E. T., & Simon, R. 2008, ApJ, 680, 349

Jackson, J. M., Rathborne, J. M., Shah, R. Y., et al. 2006, ApJS, 163, 145

Jackson, J. M., Rathborne, J. M., Foster, J. B., et al. 2013, PASA, 30, 57

Johnstone, D., Fiege, J. D., Redman, R. O., Feldman, P. A., & Carey, S. J. 2003, ApJ, 588, L37

Jurić, M., Ivezić, Ž., Brooks, A., et al. 2008, ApJ, 673, 864

Kamenetzky, J., Rangwala, N., Glenn, J., Maloney, P. R., & Conley, A. J. 2014, ApJ, submitted

Kennicutt, R. C., & Evans, N. J. 2012, ARA&A, 50, 531

Kennicutt, R. C., Calzetti, D., Aniano, G., et al. 2011, PASP, 123, 1347

Kennicutt, Jr., R. C. 1998, ApJ, 498, 541

Kennicutt, Jr., R. C., Armus, L., Bendo, G., et al. 2003, PASP, 115, 928

Kim, J.-h., Abel, T., Agertz, O., et al. 2014, ApJS, 210, 14

Kritsuk, A. G., Lee, C. T., & Norman, M. L. 2013, MNRAS, 436, 3247

Kroupa, P. 2001, MNRAS, 322, 231

Lada, C. J., & Lada, E. A. 2003, ARA&A, 41, 57

Landsman, W. B. 1995, in Astronomical Society of the Pacific Conference Series, Vol. 77, Astronomical Data Analysis Software and Systems IV, ed. R. A. Shaw, H. E. Payne, & J. J. E. Hayes, 437

Larson, R. B. 1981, MNRAS, 194, 809

Li, D., & Goldsmith, P. F. 2003, ApJ, 585, 823

Lombardi, M., & Alves, J. 2001, A&A, 377, 1023

Lombardi, M., Alves, J., & Lada, C. J. 2010, A&A, 519, L7

Lowell, P. 1908, Mars as the Abode of Life (The Macmillan company)

Marshall, D. J., Joncas, G., & Jones, A. P. 2009, ApJ, 706, 727

Marshall, D. J., Robin, A. C., Reylé, C., Schultheis, M., & Picaud, S. 2006, A&A, 453, 635

Martin, P. G., Roy, A., Bontemps, S., et al. 2012, ApJ, 751, 28

McClure-Griffiths, N. M., Dickey, J. M., Gaensler, B. M., et al. 2005, ApJS, 158, 178

McKee, C. F., & Ostriker, E. C. 2007, ARA&A, 45, 565

McMillan, P. J., & Binney, J. J. 2010, MNRAS, 402, 934

Merello, M., Evans, II, N. J., Shirley, Y. L., et al. 2014, ApJS, submitted

Molinari, S., Schisano, E., Faustini, F., et al. 2011, A&A, 530, A133

Molinari, S., Swinyard, B., Bally, J., et al. 2010a, A&A, 518, L100

—. 2010b, PASP, 122, 314

Muller, C. A., & Oort, J. H. 1951, Nature, 168, 357

Netterfield, C. B., Ade, P. A. R., Bock, J. J., et al. 2009, ApJ, 707, 1824

Nguyen Luong, Q., Motte, F., Schuller, F., et al. 2011, A&A, 529, A41

Niinuma, K., Nagayama, T., Hirota, T., et al. 2011, PASJ, 63, 9

Offner, S. S. R., Clark, P. C., Hennebelle, P., et al. 2013, ArXiv e-prints, arXiv:1312.5326

Olmi, L., Anglés-Alcázar, D., Elia, D., et al. 2013, A&A, 551, A111

Oort, J. H., Kerr, F. J., & Westerhout, G. 1958, MNRAS, 118, 379

Ossenkopf, V., & Henning, T. 1994, A&A, 291, 943

Padoan, P., Juvela, M., Goodman, A. A., & Nordlund, Å. 2001, ApJ, 553, 227

Padoan, P., & Nordlund, Å. 2002, ApJ, 576, 870

Padoan, P., Nordlund, A., & Jones, B. J. T. 1997, MNRAS, 288, 145

Perault, M., Omont, A., Simon, G., et al. 1996, A&A, 315, L165

Peretto, N., & Fuller, G. A. 2009, A&A, 505, 405, (PF09)

—. 2010, ApJ, 723, 555

Peretto, N., Fuller, G. A., Plume, R., et al. 2010, A&A, 518, L98

Persic, M., Salucci, P., & Stel, F. 1996, MNRAS, 281, 27

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. 2007, Numerical Recipes 3rd Edition: The Art of Scientific Computing, 3rd edn. (New York, NY, USA: Cambridge University Press)

Purcell, C. R., Longmore, S. N., Walsh, A. J., et al. 2012, MNRAS, 426, 1972

Rathborne, J. M., Jackson, J. M., & Simon, R. 2006, ApJ, 641, 389

Rathborne, J. M., Johnson, A. M., Jackson, J. M., Shah, R. Y., & Simon, R. 2009, ApJS, 182, 131

Reach, W. T., Megeath, S. T., Cohen, M., et al. 2005, PASP, 117, 978

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2009a, ApJ, 693, 397

Reid, M. J., Menten, K. M., Zheng, X. W., et al. 2009b, ApJ, 700, 137

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, ApJ, 783, 130

Ridge, N. A., Di Francesco, J., Kirk, H., et al. 2006, AJ, 131, 2921

Robin, A. C., Reylé, C., Derrière, S., & Picaud, S. 2003, A&A, 409, 523

Robitaille, T. P. 2011, A&A, 536, A79

Robitaille, T. P., Churchwell, E., Benjamin, R. A., et al. 2012, A&A, 545, A39, (R12)

Robitaille, T. P., Whitney, B. A., Indebetouw, R., & Wood, K. 2007, ApJS, 169, 328

Rodriguez-Fernandez, N. J., & Combes, F. 2008, A&A, 489, 115

Roman-Duval, J., Jackson, J. M., Heyer, M., et al. 2009, ApJ, 699, 1153

Roman-Duval, J., Jackson, J. M., Heyer, M., Rathborne, J., & Simon, R. 2010, ApJ, 723, 492, (RD10)

Román-Zúñiga, C. G., Lada, C. J., Muench, A., & Alves, J. F. 2007, ApJ, 664, 357

Rosolowsky, E. 2005, PASP, 117, 1403

Rosolowsky, E., Ginsburg, A., & Ellsworth-Bowers, T. 2010a, in Astronomical Society of the Pacific Conference Series, Vol. 438, The Dynamic Interstellar Medium: A Celebration of the Canadian Galactic Plane Survey, ed. R. Kothes, T. L. Landecker, & A. G. Willis, 76

Rosolowsky, E., Dunham, M. K., Ginsburg, A., et al. 2010b, ApJS, 188, 123

Rosolowsky, E. W., Pineda, J. E., Foster, J. B., et al. 2008a, ApJS, 175, 509

Rosolowsky, E. W., Pineda, J. E., Kauffmann, J., & Goodman, A. A. 2008b, ApJ, 679, 1338

Ruphy, S., Robin, A. C., Epchtein, N., et al. 1996, A&A, 313, L21

Russeil, D., Pestalozzi, M., Mottram, J. C., et al. 2011, A&A, 526, A151

Russell, H. N. 1914, Popular Astronomy, 22, 275

Rygl, K. L. J., Brunthaler, A., Reid, M. J., et al. 2010, A&A, 511, A2

Salpeter, E. E. 1955, ApJ, 121, 161

Sanders, D. B., Scoville, N. Z., & Solomon, P. M. 1985, ApJ, 289, 373

Sato, M., Reid, M. J., Brunthaler, A., & Menten, K. M. 2010, ApJ, 720, 1055

Schlingman, W. M., Shirley, Y. L., Schenk, D. E., et al. 2011, ApJS, 195, 14

Schmidt, M. 1957, Bull. Astron. Inst. Netherlands, 13, 247

Schnee, S., Enoch, M., Johnstone, D., et al. 2010, ApJ, 718, 306

Schneider, N., Csengeri, T., Bontemps, S., et al. 2010, A&A, 520, A49

Schönrich, R., Binney, J., & Dehnen, W. 2010, MNRAS, 403, 1829

Schuller, F., Menten, K. M., Contreras, Y., et al. 2009, A&A, 504, 415

Scoville, N. Z., & Solomon, P. M. 1975, ApJ, 199, L105

Sewilo, M., Watson, C., Araya, E., et al. 2004, ApJS, 154, 553

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, (S06)

Simon, R., Rathborne, J. M., Shah, R. Y., Jackson, J. M., & Chambers, E. T. 2006b, ApJ, 653, 1325

Slipher, V. M. 1914, Lowell Observatory Bulletin, 2, 66

Solomon, P. M., Rivolo, A. R., Barrett, J., & Yahil, A. 1987, ApJ, 319, 730

Stacy, J. A., & Thaddeus, P. 1991, in Astronomical Society of the Pacific Conference Series, Vol. 16, Atoms, Ions and Molecules: New Results in Spectral Line Astrophysics, ed. A. D. Haschick & P. T. P. Ho, 197

Stetson, P. B. 1987, PASP, 99, 191

Stil, J. M., Taylor, A. R., Dickey, J. M., et al. 2006, AJ, 132, 1158

Swift, J. J., & Beaumont, C. N. 2010, PASP, 122, 224

Tassis, K., Christie, D. A., Urban, A., et al. 2010, MNRAS, 408, 1089

Taylor, A. R., Gibson, S. J., Peracaula, M., et al. 2003, AJ, 125, 3145

Vázquez-Semadeni, E., Ryu, D., Passot, T., González, R. F., & Gazol, A. 2006, ApJ, 643, 245

Veltchev, T. V., Donkov, S., & Klessen, R. S. 2013, MNRAS, 432, 3495

Walsh, A. J., Breen, S. L., Britton, T., et al. 2011, MNRAS, 416, 1764

Weingartner, J. C., & Draine, B. T. 2001, ApJ, 548, 296, (WD01)

Westerhout, G. 1958, Bull. Astron. Inst. Netherlands, 14, 215

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., Reid, M. J., Menten, K. M., et al. 2009, ApJ, 693, 413

Xu, Y., Li, J. J., Reid, M. J., et al. 2013, ApJ, 769, 15

Appendix A

The `distance-omnibus` Code Base

A.1 Motivation

The motivation for this thesis was to compute distances to and physical properties of a large number of dense molecular cloud structures detected by the Bolocam Galactic Plane Survey (BGPS). While previous studies of isolated regions had done much of the distance resolution work by hand on a source-by-source basis, the large size (8,000+ objects) of the BGPS catalog makes this infeasable. An automated procedure for distance assignment is a necessity for large-scale blind surveys of the Galactic plane.

The basic idea for the DPDF formalism predates this thesis, as do some of the basic code blocks incorporated into the `distance-omnibus` framework. The Bayesian DPDF framework itself, whereby the probability of an object being at various distances is computed, requires a series of interconnected modules that compute the kinematic distance likelihood function and prior DPDFs. Extending this computational requirement to be completely automated was necessary in light of the large number of objects to be studied. In addition to computing posterior DPDFs in an automated fashion, the `distance-omnibus` code base also autonomously computes the distance statistics for each object (**i.e.**, KDA resolution, confidence in that resolution, single-value distance estimates with uncertainty).

The original version of `distance-omnibus` was written specifically for the BGPS, version 1, culminating in the publication of Ellsworth-Bowers et al. (2013). At about that time, the version 2 release was being prepared, and it was decided to utilize the new data for any following distance work, especially given the improvements in source flux-density recovery and noise properties in the maps (not to mention the correction of the calibration error in the V1 data). Rather than simply change all the hardwired-values in the code to work with BGPS V2, the framework was changed to be usable with any (sub-)millimeter continuum survey, greatly enhancing the value of the work.

A.2 Modular Framework

The `distance-omnibus` code base is written in the Interactive Data Language (IDL),[1] but could (should) be ported to a more universally-available (read: "free") language like `python`. It relies upon several external libraries, inlcuding the `idlastro` library[2] of astronomy-related routines, the Coyote Graphics Library[3] of device-independent plotting routines, and the Markwardt library[4] containing curve-fitting and optimization routines.

Like the DPDF framework itself, the `distance-omnibus` code base is constructed in a modular fashion to allow for the use of multiple kinematic distance likelihood functions and the application of any number of prior DPDFs for any object. The driving routine, `distance_omnibus.pro`, calls individual DPDF routines (all called `prob_*.pro`) based on entries in the `dpdf_params.conf` configuration file; different sets or orders of DPDFs may be run without modifying or recompiling the code. Furthermore, information about the dust-continuum survey being used (**e.g.**, central frequency, chosen dust opacity, beam size, etc.) is stored in a different configuration file, and distances may be computed for different surveys, again without modifying or recompiling the code.

The basic operation of `distance-omnibus` requires a collection of locally created files constructed from the survey data and catalog. All of these files are created automatically by the driving routine if they do not already exist. Additionally, some prior DPDFs (such as EMAFs) require specifically-processed versions of ancillary data. Almost all of the custom-processed files are generated automatically – the sole exception is listed in the documentation and requires substantial

[1] `http://www.exelisvis.com/ProductsServices/IDL.aspx`
[2] `http://idlastro.gsfc.nasa.gov`
[3] `http://www.idlcoyote.com/graphics_tips/coyote_graphics.php`
[4] `http://www.physics.wisc.edu/~craigm/idl/idl.html`

computational power to produce.[5] Furthermore, the code automatically generates a structure containing the combined velocity data specified by the user for the computation of kinematic distance likelihood functions.

A.3 Public Access and Future Development

Since April 2014, `distance-omnibus` has been available on GitHub,[6] a web-based code hosting service using the `git` revision control system. Use of `git` and GitHub allows for distributed development with a free, publicly-available central repository, making future extensions of `distance-omnibus` not dependent upon on the attention of a single individual. With the publication of Ellsworth-Bowers et al. (2014) in ApJ, version 1.0 will be released for public use.[7] Further development and maintenance will be undertaken by other members of the BGPS team or larger astronomical community.

The modular nature of the code means new DPDF methods (like those discussed in §5.2.1) may be developed and implemented through adding the method name to the `dpdf_params.conf` configuration file. As new ancillary data sets and KDA resolution techniques become available, `distance-omnibus` has the capacity to easily incorporate "everything but the kitchen sink" in its automated distance estimation framework for dense molecular cloud structures.

[5] These are the star-subtracted versions of the **Spitzer**/GLIMPSE $\lambda = 8\ \mu$m images; computation may take up to several days, depending on available computing power. If these images are not present, an error message will be displayed and the user is instructed how to produce them.

[6] `https://github.com`

[7] `https://github.com/BGPS/distance-omnibus`