



# Revisiting the Vertical Distribution of HI Absorbing Clouds in the Solar Neighborhood

Trey V. Wenger<sup>1</sup> , Daniel R. Rybarczyk<sup>1</sup> , and Snežana Stanimirović<sup>2</sup>

<sup>1</sup> NSF Astronomy & Astrophysics Postdoctoral Fellow, Department of Astronomy, University of Wisconsin–Madison, Madison, WI 53706, USA; [twenger2@wisc.edu](mailto:twenger2@wisc.edu)

<sup>2</sup> Department of Astronomy, University of Wisconsin–Madison, Madison, WI 53706, USA

Received 2024 March 14; revised 2024 March 27; accepted 2024 March 27; published 2024 May 9

## Abstract

The vertical distribution of cold neutral hydrogen (HI) clouds is a constraint on models of the structure, dynamics, and hydrostatic balance of the interstellar medium. In 1978, Crovisier pioneered a method to infer the vertical distribution of HI absorbing clouds in the solar neighborhood. Using data from the Nançay 21 cm absorption survey, Crovisier determined the mean vertical displacement of cold HI clouds,  $\langle |z| \rangle$ . We revisit that author's analysis and explore the consequences of truncating the HI absorption sample in Galactic latitude. For any nonzero latitude limit, we find that the quantity inferred by Crovisier is not the mean vertical displacement but rather a ratio involving higher moments of the vertical distribution. The resultant distribution scale heights are thus  $\sim 1.5$  to  $\sim 3$  times smaller than previously determined. In light of this discovery, we develop a Bayesian Monte Carlo Markov Chain method to infer the vertical distribution of HI absorbing clouds. We fit our model to the original Nançay data and find a vertical distribution moment ratio  $\langle |z|^3 \rangle / \langle |z|^2 \rangle = 97 \pm 15$  pc, which corresponds to a Gaussian scale height  $\sigma_z = 61 \pm 9$  pc, an exponential scale height  $\lambda_z = 32 \pm 5$  pc, and a rectangular half-width  $W_{z,1/2} = 129 \pm 20$  pc. Consistent with recent simulations, the vertical scale height of cold HI clouds appears to remain constant between the inner Galaxy and the Galactocentric distance of the solar neighborhood. Local fluctuations might explain the large-scale height observed at the same Galactocentric distance on the far side of the Galaxy.

*Unified Astronomy Thesaurus concepts:* Interstellar atomic gas (833); Neutral hydrogen clouds (1099); Interstellar dynamics (839); Cold neutral medium (266); Radio astronomy (1338); Milky Way disk (1050)

*Supporting material:* machine-readable table

## 1. Introduction

The structure of the interstellar medium (ISM) is determined by the complex interplay of many physical processes (e.g., McKee & Ostriker 1977). In particular, the vertical distribution (i.e., perpendicular to the Galactic plane) of interstellar matter is set by the local gravitational potential, energy injection by stellar feedback and supernovae, the radiation field, and the physical conditions (e.g., pressure, turbulence, and magnetic support) of the ISM (e.g., Kim et al. 2010; Ostriker & Kim 2022). Characterizing the vertical distribution of the various phases of the ISM ultimately constrains models and simulations of ISM physics, star formation, and galactic evolution (e.g., Hopkins et al. 2018).

Neutral atomic hydrogen (HI) is pervasive in the ISM and exists in pressure equilibrium in two stable phases: the cold neutral medium (CNM) and the warm neutral medium (WNM; Field et al. 1969; McKee & Ostriker 1977; Wolfire et al. 2003). It is within clumps of CNM gas that molecular clouds, and ultimately stars, form (McKee & Ostriker 2007). The WNM is easily traced via the 21 cm hyperfine transition of HI in emission, and its vertical distribution has been extensively studied. Levine et al. (2006), for example, find a strong anticorrelation between the vertical scale height of HI emission and spiral structure inferred from the HI surface density. This suggests that the gravitational potential of spiral arms acts to squeeze the WNM (e.g., Kalberla & Kerp 2009). Given its low kinetic and spin temperature ( $T \sim 100$  K), the CNM is most

easily observed via 21 cm absorption toward background continuum sources (e.g., Dickey et al. 1983) or in self-absorption toward background HI emission (e.g., Kavars et al. 2003). Such HI absorption observations are challenging to interpret in the Milky Way ISM, due to the blending of HI features in both emission and absorption along sight lines through the Galactic midplane (e.g., Dickey et al. 2003; McClure-Griffiths et al. 2023 and references therein). Nonetheless, several studies have explored the differences between the vertical HI distribution as traced by HI emission and absorption. Dickey et al. (2009), for example, explore the Galactocentric radial and azimuthal variation in HI scale height in the outer Galaxy. They find similar scale heights between the two HI phases.

In the inner Galaxy, Dickey et al. (2022) use HI absorption observations from the Galactic Australia Square Kilometer Array Pathfinder (GASKAP) pilot survey to constrain the vertical scale height of CNM clouds. The complications of interpreting HI absorption spectra in the inner Galaxy are mitigated by considering only those components near the tangent point, where distance ambiguities are minimized. They find a compact vertical CNM cloud distribution, with an inferred scale height  $\sigma_z \simeq 50\text{--}90$  pc. This distribution closely matches the scale height of the thin molecular cloud layer in the inner Galaxy,  $\sigma_z \simeq 40\text{--}60$  pc (Bronfman et al. 1988; Heyer & Dame 2015; Su et al. 2021). Furthermore, Dickey et al. (2022) infer a scale height of  $\sigma_z \simeq 200\text{--}700$  pc for HI absorbing clouds beyond the solar circle in the fourth Galactic quadrant, which suggests significant flaring of the CNM in the Galactic outskirts. This is consistent with Galactic simulations where the HI disk becomes more vertically extended at the edges of galaxies due to the decreased pressure of the ISM, although the



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](#). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

effect is seen only in the warm H I and not the CNM (Smith et al. 2023).

Dickey et al. (2022) also estimate the vertical scale height of CNM clouds near the solar orbit by considering those H I absorption detections with velocities relative to the local standard of rest (LSR) near  $0 \text{ km s}^{-1}$ . At these velocities, their sample includes both local H I absorption as well as H I absorption near the solar Galactocentric distance,  $R_0$ , along a Galactic longitude  $\ell \simeq 340^\circ$ , which, for  $R_0 = 8.34 \text{ kpc}$  (Reid et al. 2014), corresponds to a heliocentric distance  $d \simeq 15.6 \text{ kpc}$ . They find that the latitude distribution of these H I absorption features is a blend of two Gaussian distributions with different widths, presumably corresponding to the local and distant CNM clouds. Considering only the narrow component, they estimate a vertical scale height of  $\sigma_z \simeq 160 \text{ pc}$  for the CNM at the solar orbit.

There are few experiments investigating the vertical distribution of the CNM in the solar neighborhood. In fact, nearly every reference that we could find in the literature ultimately pointed back to a single experiment: Crovisier (1978). In this work, the author develops a statistical model to infer the vertical scale height of local H I absorbing clouds. Using data from the Nançay 21 cm absorption survey (Crovisier et al. 1978), Crovisier (1978) constrains the mean absolute vertical displacement of CNM clouds in the solar neighborhood as  $\langle |z| \rangle = 107 \pm 29 \text{ pc}$ . That author claims that the data are inconsistent with an exponential vertical distribution, and the inferred  $\langle |z| \rangle$  implies a scale height  $\sigma_z = 134 \pm 36 \text{ pc}$  for a Gaussian vertical distribution. Belfort & Crovisier (1984) expand this experiment to include more H I absorption data, but the estimated mean absolute vertical displacement is similar:  $\langle |z| \rangle = 92 \pm 12 \text{ pc}$  for clouds with  $5^\circ < |b| < 30^\circ$ . These results are consistent with the Dickey et al. (2022) inference at  $R_0$  in the fourth Galactic quadrant. Altogether, we begin to develop a clear picture of a vertical CNM distribution that is thin in the inner Galaxy, mimicking the molecular gas distribution, and slowly flares with Galactocentric distance (e.g., see Section 6.4.2. and Figure 10 in McClure-Griffiths et al. 2023).

In this work, we revisit the Crovisier (1978) analysis in preparation for its application to the ever-growing sample of H I absorption spectra (e.g., HIBIGCAT; McClure-Griffiths et al. 2023) and next-generation H I surveys (e.g., GASKAP; Dickey et al. 2013). In particular, we discover a statistical bias that alters the interpretation of the Crovisier (1978) result, which we demonstrate through both a simple simulation as well as a robust derivation. We develop a Bayesian model that performs a similar inference as in the original work, and using the original Nançay data, we implement a Monte Carlo Markov Chain (MCMC) method to infer the vertical distribution of CNM clouds in the solar neighborhood. Finally, we discuss the implications of these results in light of modern Galactic ISM simulations.

## 2. Method

Our goal is to infer the shape of the vertical distribution of H I absorbing clouds (or any other tracer) using only their sky positions and radial velocities. Crovisier (1978) introduces a statistical method to perform this inference, which we briefly summarize here (see their Section 2 for more details). For convenience and simplicity, we adopt a slightly different notation:  $d$  is the heliocentric distance and  $z \equiv |z|$  is the absolute vertical displacement of the cloud from the Galactic

midplane. We assume that the Sun is located at  $z = 0$  and the midplane is defined by Galactic latitude  $b = 0^\circ$ .

The observed radial velocity of a cloud is related to its distance due to Galactic rotation. For a cloud located at a given Galactic longitude and latitude,  $(\ell, b)$ , the radial velocity in the LSR frame is

$$V_{\text{LSR}} = V_{\odot, \text{LSR}}(\ell, b) + V_{\text{rot}}(d, \ell, b) + V_t, \quad (1)$$

where  $V_{\odot, \text{LSR}}$  is the solar motion with respect to the LSR projected onto the line of sight,  $V_{\text{rot}}$  is the cloud's motion due to Galactic rotation projected onto the line of sight, and  $V_t$  is the random velocity of the cloud along the line of sight. Crovisier (1978) uses a local approximation for Galactic rotation defined by the first Oort constant, but any model that predicts  $V_{\text{rot}}$  for a given distance and sky position is appropriate.

Given that radial velocity is only a weak indicator of distance in the solar neighborhood,  $d$  cannot be determined for any individual cloud with sufficient precision to infer the shape of the vertical distribution of clouds. If we assume a plane-parallel model for clouds in the solar neighborhood, then, Crovisier (1978) argues, we can relate the first moment of a cloud's distance probability distribution (i.e., the expectation value),  $\langle d \rangle$ , to the first moment of the vertical distribution of clouds (i.e., the mean absolute vertical displacement),  $\langle |z| \rangle$ , by

$$\langle d \rangle = \frac{\langle |z| \rangle}{\sin|b|}. \quad (2)$$

With a sufficiently large sample of clouds, the central limit theorem suggests that we can infer  $\langle |z| \rangle$  from these distance point estimates despite their imprecision.

Finally, we assume that  $V_t$  is truly random; that is, we assume that there are no systematic departures from the Galactic rotation model in our sample of clouds. Crovisier (1978) uses a least-squares algorithm to minimize

$$\sum V_t^2 = \sum [V_{\text{LSR}} - V_{\odot, \text{LSR}} - V_{\text{rot}}(\langle d \rangle)]^2, \quad (3)$$

where the sum is taken over all of the clouds, in order to infer the mean absolute vertical displacement,  $\langle |z| \rangle$ , the rms of  $V_t$ , and the parameters that define the solar noncircular motion ( $V_{\odot, \text{LSR}}$ ) and Galactic rotation ( $V_{\text{rot}}$ ).

### 2.1. The Bias

We must first derive Equation (2), which relates the first moment of a cloud's distance probability distribution to the first moment of the vertical distribution of clouds, before we can demonstrate the bias introduced by a subtle assumption. Consider a population of clouds that follows an absolute vertical displacement distribution defined by the probability density  $P_Z(z)$ . We adopt a notation where the arguments of  $P$  represent random variables and the subscripts of  $P$  represent the space over which the probability is defined. In this case,  $z \in Z = \mathbb{R}^+$ . Assuming that the clouds are distributed uniformly across the Galactic plane (i.e., a plane-parallel model), then the distance probability density for a cloud with Galactic latitude  $b$  is

$$P_D(d) = P_Z(z) \frac{dz}{dd} = P_Z(z) \sin|b|, \quad (4)$$

where  $z = d \sin|b|$ . Thus,  $P_Z(z) = P_D(d)/\sin|b|$  for  $\sin|b| > 0$ . The first moment of the distance probability distribution is

$$\begin{aligned}\langle d \rangle &= \int_0^\infty dP_D(d)dd \\ &= \frac{1}{\sin|b|} \int_0^\infty zP_Z(z)dz \\ &= \frac{\langle |z| \rangle}{\sin|b|},\end{aligned}\quad (5)$$

for  $\sin|b| > 0$ , where  $\langle |z| \rangle$  is the first moment of the absolute vertical displacement distribution. Presumably, this is how Crovisier (1978) arrives at Equation (2).

A simple simulation reveals a problem with Equation (2) when we truncate the data in Galactic latitude. Crovisier (1978) motivates the need for latitude truncation in order to (1) restrict the sample to local HI clouds and (2) remove complex HI absorption spectra in the Galactic plane. We demonstrate the problem by drawing a random sample of clouds from a uniform distribution in the Galactic plane out to some maximum midplane distance,  $R_{\max}$ , such that  $(x, y) \sim U(-R_{\max}, R_{\max})$ , and a half-normal distribution in absolute vertical displacement with some variance,  $\sigma_z^2$ , such that  $z \sim N_{1/2}(\sigma_z^2)$ . For a half-normal distribution, the standard deviation is related to the first moment by  $\langle |z| \rangle = \sigma_z \sqrt{2/\pi}$ . The midplane distance is  $r = \sqrt{x^2 + y^2}$  and the Galactic latitude is  $b = \tan^{-1}(z/r)$ . For a random sample of 1,000,000 clouds with  $R_{\max} = 10$  kpc and  $\sigma_z = 100$  pc, we use Equation (2) to calculate  $\langle |z| \rangle = \langle d \sin|b| \rangle \simeq 80$  pc, which is consistent with the expectation for a half-normal vertical distribution,  $\langle |z| \rangle = \sigma_z \sqrt{2/\pi} \simeq 80$  pc. If we remove the clouds with  $|b| < b_{\min}$ , however, then the discrepancy emerges. For the same simulated data with  $b_{\min} = 1^\circ$ , for example, we find  $\langle d \sin|b| \rangle \simeq 120$  pc. For  $b_{\min} \geq 2^\circ$ , we find  $\langle d \sin|b| \rangle \simeq 160$  pc. There is a factor-of-two difference from the Equation (2) prediction that appears to be independent of  $b_{\min}$  for  $b_{\min} \geq 2^\circ$  (i.e., for  $R_{\max} \gg \sigma_z/\tan b_{\min}$ ). Figure 1 demonstrates the result of this simple simulation.

## 2.2. The Correction

Equation (2) is only correct for  $b_{\min} = 0^\circ$ . For any  $b_{\min} > 0^\circ$ , we must derive the distance expectation value while properly considering the joint dependency between  $d$ ,  $z$ , and  $b$ . Consider again a plane-parallel geometry where clouds are distributed uniformly in the Galactic midplane out to a maximum midplane distance  $R_{\max}$  and follow some absolute vertical displacement distribution  $P_Z(z)$ . The joint probability distribution for a cloud's location in the heliocentric Galactic Cartesian frame is

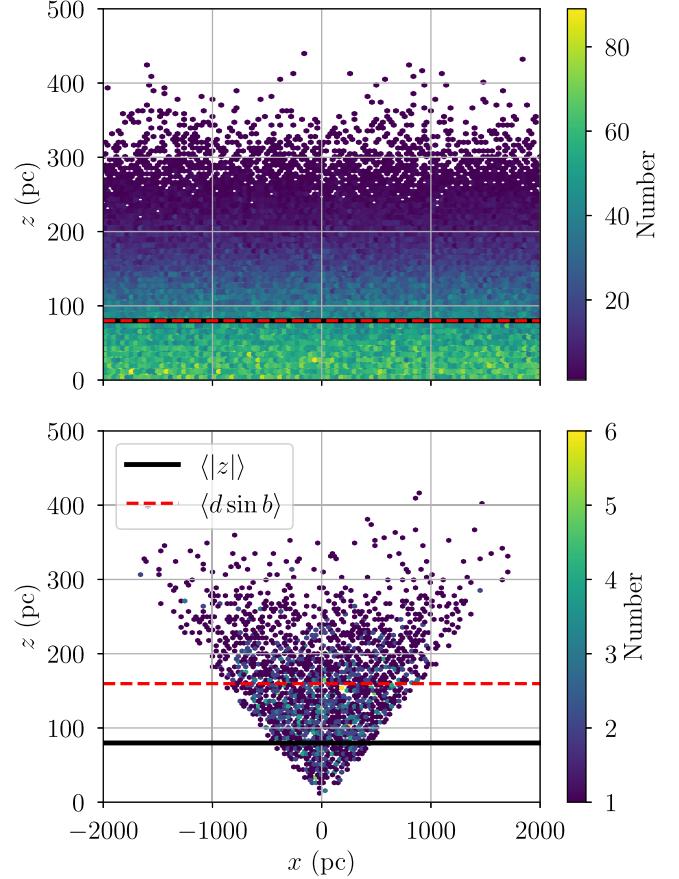
$$\begin{aligned}P_{XYZ}(x, y, z) &= P_X(x)P_Y(y)P_Z(z) \\ &= \frac{1}{\pi R_{\max}^2} P_Z(z).\end{aligned}\quad (6)$$

The transformation to the cylindrical frame is

$$P_{RLZ}(r, \ell, z) = \frac{r}{\pi R_{\max}^2} P_Z(z),\quad (7)$$

and that to the spherical frame is

$$P_{DLB}(d, \ell, b) = \frac{d^2 \cos^3 b}{\pi R_{\max}^2} P_Z(d \sin|b|).\quad (8)$$



**Figure 1.** The binned heliocentric  $(x, z)$  distribution of simulated clouds drawn from a uniform distribution in  $x$  and  $y$  and a half-normal distribution in  $z$  with standard deviation  $\sigma_z = 100$  pc. Note that the scale of the  $x$  axis is much larger than that of the  $z$  axis. The top panel includes all clouds whereas the bottom panel only includes clouds with  $b > b_{\min} = 10^\circ$ . The solid black lines indicate  $\langle |z| \rangle = \sigma_z \sqrt{2/\pi} \simeq 80$  pc for a half-normal distribution. The dashed red lines indicate  $\langle d \sin|b| \rangle$  derived from the simulated clouds. Clearly,  $\langle |z| \rangle \neq \langle d \sin|b| \rangle$  for  $b_{\min} > 0^\circ$ .

Marginalized over distance, the probability density is

$$\begin{aligned}P_{LB}(\ell, b) &= \int_0^{R_{\max}/\cos b} P_{DLB}(d, \ell, b)dd \\ &= \int_0^{R_{\max} \tan|b|} \frac{z^2 \cos b}{\pi R_{\max}^2 \sin^3|b|} P_Z(z)dz.\end{aligned}\quad (9)$$

If we require that  $P_Z(z) \rightarrow 0$  as  $z \rightarrow R_{\max} \tan|b|$  (i.e., the artificially truncated midplane distribution does not appreciably truncate the  $z$  distribution along the line of sight toward  $b$ ), then

$$P_{LB}(\ell, b) \simeq \frac{\cos b}{\pi R_{\max}^2 \sin^3|b|} \langle |z|^2 \rangle,\quad (10)$$

where  $\langle |z|^2 \rangle = \int_0^\infty z^2 P_Z(z)dz$  is the second moment of the  $z$  distribution. The conditional probability distribution—the probability of finding a cloud at a given distance along a given line of sight—is

$$\begin{aligned}P_{D|LB}(d|\ell, b) &= \frac{P_{DLB}(d, \ell, b)}{P_{LB}(\ell, b)} \\ &= \frac{d^2 \sin^3|b|}{\langle |z|^2 \rangle} P_Z(d \sin|b|).\end{aligned}\quad (11)$$

The first moment of this distribution is

$$\langle d|\ell, b \rangle = \int_0^{R_{\max}/\cos b} dP_{D|LB}(d|\ell, b) dd. \quad (12)$$

If we again require that  $P_{D|LB}(d|\ell, b) \rightarrow 0$  as  $d \rightarrow R_{\max}/\cos b$  (i.e., the artificially truncated midplane distribution does not appreciably truncate the distance distribution along the line of sight toward  $b$ ), then

$$\langle d|\ell, b \rangle \simeq \frac{\langle |z|^3 \rangle}{\langle |z|^2 \rangle \sin|b|}. \quad (13)$$

Thus, we find that, for any  $b_{\min} > 0^\circ$ , the distance expectation value is not directly related to the first moment of the absolute vertical displacement distribution,  $\langle |z| \rangle$ , as in Equation (2), but rather the ratio of the third moment (skewness) to the second moment (variance),  $\langle |z|^3 \rangle / \langle |z|^2 \rangle$ .

Equipped with this corrected derivation, we can test the result against our simple simulation. For a half-normal distribution,  $\langle |z| \rangle = \sigma_z \sqrt{2/\pi}$ ,  $\langle |z|^2 \rangle = \sigma_z^2$ , and  $\langle |z|^3 \rangle = 2\sigma_z^3 \sqrt{2/\pi}$ . For our simulation parameters  $R_{\max} = 10$  kpc and  $\sigma_z = 100$  pc, the assumptions of our derivation are valid for  $b_{\min} \gtrsim 2^\circ$ , since  $P_z(R_{\max} \tan b_{\min}) \simeq 0$  and  $P_{D|LB}(R_{\max}/\cos b_{\min}|\ell, b_{\min}) \simeq 0$ . The moment ratio is related to the standard deviation for a half-normal distribution by

$$\langle d \sin|b| \rangle \simeq \frac{\langle |z|^3 \rangle}{\langle |z|^2 \rangle} = 2\langle |z| \rangle = 2\sqrt{\frac{2}{\pi}}\sigma_z. \quad (14)$$

Here, we see the erroneous factor of two. For our simulation parameters, we predict  $\langle d \sin|b| \rangle \simeq 160$  pc, which is consistent with our simulation results.

For the other vertical distribution shapes considered by Crovisier (1978), we calculate the factor by which the author overestimates the distribution width due to latitude truncation, which is simply the ratio  $(\langle |z|^3 \rangle / \langle |z|^2 \rangle) / \langle |z| \rangle$ . As demonstrated for the half-normal distribution, this factor is 2. For an exponential distribution with scale height  $\lambda_z$ , the moments are  $\langle |z|^n \rangle = n! \lambda_z^n$ , and thus the factor is 3. Finally, for a rectangular distribution parameterized by the half-width,  $W_{z,1/2}$ , the moments are  $\langle |z|^n \rangle = W_{z,1/2}^n / (n+1)$ , so the factor is 1.5. Thus, for a Gaussian, exponential, or rectangular distribution, Crovisier (1978) overestimates the distribution width by a factor of 2, 3, or 1.5, respectively, due to the subtle bias introduced by latitude truncation.

### 3. Data and Analysis

Here, we use the original Crovisier et al. (1978) data to reanalyze the vertical distribution of H I absorbing clouds in the solar neighborhood in light of the aforementioned latitude truncation bias. Our objectives are to (1) reproduce the Crovisier (1978) original results, (2) perform a similar least-squares analysis using a modern Galactic rotation model, and (3) infer the vertical cloud distribution using a robust, outlier-tolerant Bayesian model that incorporates the uncertainties in the assumed Galactic rotation model.

#### 3.1. Data

The Crovisier (1978) data come from the Nançay 21 cm absorption survey (Crovisier et al. 1978). At 21 cm, the Nançay radio telescope has a half-power beam with of  $3' 1$  in R.A. and  $2' 1$  in decl. The survey results contain H I absorption spectra

**Table 1**  
H I Absorption Sample

Source Name	$\ell$ (deg)	$b$ (deg)	$V_{\text{LSR}}$ (km s $^{-1}$ )
0007+12	107.51	-48.89	-8.1 $\pm$ 1.1
0013+79	121.28	16.54	1.4 $\pm$ 0.6
0026+34	117.86	-27.71	-3.3 $\pm$ 0.4
0038+32	120.47	-29.65	-1.2 $\pm$ 1.2
0038+09	118.72	-52.73	-7.8 $\pm$ 0.2

**Note.**

<sup>a</sup> These absorption parameters were determined “by eye” by Crovisier et al. (1978), so we assign a LSR velocity uncertainty of 1.0 km s $^{-1}$ .

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

toward 819 extragalactic continuum sources with a spectral resolution of 6 kHz ( $\sim 1.25$  km s $^{-1}$ ). Unfortunately, we were unable to find these data in a machine-readable format, so we transcribed the data by hand from Table 2 in Crovisier et al. (1978). Following the selection criteria in Crovisier (1978), we only copied rows for which the Galactic latitude is  $|b| > 10^\circ$  and the absorption detection is not marked as “dubious” (remark “P” in their Table 2). The source names, Galactic coordinates, and LSR velocities for these 297 absorption features are listed in Table 1. Crovisier (1978) claims to have 299 absorption detections meeting these criteria, but we are unable to reproduce this number. The absorption parameters for three detections (two toward 0538+49 and one toward 0725+14) were determined “by eye” by Crovisier et al. (1978), and thus do not have an associated velocity uncertainty. Figure 2 shows the Galactic longitude, latitude, and LSR velocity distributions for this sample. There are minor differences between our LSR velocity distribution and the similar distribution in Crovisier (1978; the top panel of their Figure 1), which we attribute to the slightly different sample size of our transcribed data.

#### 3.2. Least Squares

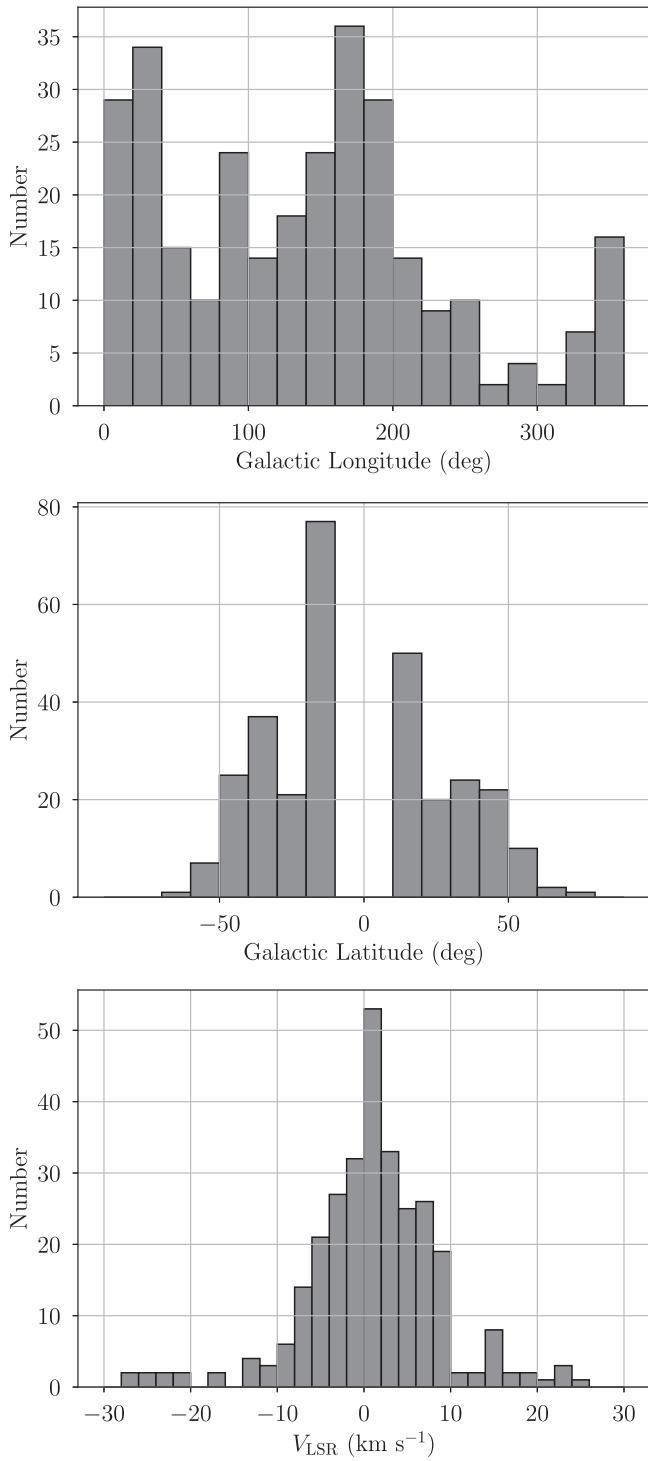
We begin our analysis using the least-squares method described by Crovisier (1978) in order to reproduce that author’s results and verify the accuracy of the transcribed data. Following Crovisier (1978), we adopt a local approximation for Galactic rotation defined in terms of Oort’s  $A$  constant,

$$V_{\text{rot}}(d) = Ad \sin[2(\ell - \ell_0)] \cos^2 b, \quad (15)$$

where  $\ell_0$  is the nodal deviation and  $A = 15$  km s $^{-1}$  kpc $^{-1}$ . The solar motion with respect to the LSR is defined as

$$V_{\odot, \text{LSR}} = (U_{\odot} \cos \ell + V_{\odot} \sin \ell) \cos b + W_{\odot} \sin b, \quad (16)$$

where  $U_{\odot}$ ,  $V_{\odot}$ , and  $W_{\odot}$  are the solar motion velocities in the direction of the Galactic center, solar orbit, and north Galactic pole, respectively. We assign each cloud to its distance expectation value (Equation (13)) given the  $z$  distribution moment ratio,  $\langle |z|^3 \rangle / \langle |z|^2 \rangle$ , and we use a least-squares algorithm to search for the solar motion velocities,  $z$  distribution moment ratio, and nodal deviation that minimize the velocity residuals (Equation (3)). Note that, following Crovisier (1978), we ignore the LSR velocity uncertainties of individual H I absorption detections, since the velocity residuals are dominated by the random component,  $V_r$ . The results of this analysis, applied to the same Galactic latitude-selected and LSR velocity-selected



**Figure 2.** Galactic longitude (top), latitude (middle), and LSR velocity,  $V_{\text{LSR}}$ , (bottom) distributions of the Crovisier et al. (1978) H I absorption detections. Three detections with  $V_{\text{LSR}} < -40 \text{ km s}^{-1}$  are excluded from the bottom panel for clarity.

subsets of data listed in Tables 1 and 2 of Crovisier (1978), are presented in Table 2. For each subset, we list the number of absorption features,  $N$ , optimized model parameters,  $U_{\odot}$ ,  $V_{\odot}$ ,  $W_{\odot}$ ,  $\ell_0$ , and  $\langle|z|^3\rangle/\langle|z|^2\rangle$ , and root mean square residual velocity,  $\sigma_V$ .

As expected, the results of our reproduction of the Crovisier (1978) least-squares analysis are nearly identical to the original

results, except that we now correctly interpret the inferred quantity as  $\langle|z|^3\rangle/\langle|z|^2\rangle$  rather than  $\langle|z|\rangle$ . This similarity suggests that we have accurately transcribed the original data and have implemented the least-squares analysis correctly. We are unable to rectify the minor differences in sample size, and we attribute the small differences in inferred values to the slightly different samples.

### 3.3. Updated Least Squares

A modern reanalysis of the Crovisier (1978) data is warranted in order to quantify the effect of an updated Galactic rotation model. Here, we adopt the A5 model of Reid et al. (2019), which is an axisymmetric Galactic rotation curve fit to the positions and kinematics of masers associated with high-mass star-forming regions. Since our sample of clouds is limited to the solar neighborhood, the most important difference between this model and the local approximation used in the previous analysis is the shape of the rotation curve at the solar Galactocentric distance. This shape is parameterized by Oort's  $A$  constant, which, for the Reid et al. (2019) model, is  $A \simeq 15.2 \text{ km s}^{-1} \text{ kpc}^{-1}$  and not so different from the previous assumption. Nonetheless, by incorporating this nonlocal Galactic rotation model, we can be confident that second-order effects for potentially distant H I absorbing clouds are properly considered.

We use another least-squares analysis to infer the model parameters by minimizing the velocity residuals given by Equation (3) and again ignoring the LSR velocity uncertainties of individual clouds. We fix  $R_0$  and the Galactic rotation model parameters  $a_2$  and  $a_3$  to the mean point estimates from the Reid et al. (2019) A5 model but allow the parameters defining the solar motion with respect to the LSR to vary. This least-squares method has one fewer free parameter (the nodal deviation) than the original Crovisier (1978) analysis. Table 2 shows the results of the least-squares analysis using the Reid et al. (2019) Galactic rotation model for the same subsets of data. Unsurprisingly, since the local shape of the Galactic rotation curve is similar between both analyses, the inferred distribution of H I absorbing clouds is nearly the same between the two methods.

### 3.4. MCMC Moment Ratio

There are two important deficiencies in the aforementioned least-squares analyses: outliers and uncertainties in the Galactic rotation model. As noted by Crovisier (1978) and demonstrated in Table 2, the inferred shape of the H I absorbing cloud distribution strongly depends on the LSR velocities included in the sample. This dependency suggests that the data are not sufficiently described by the model; outliers bias the inference. Furthermore, we use point estimates for the Galactic rotation model parameters (i.e., Oort's  $A$  constant as in Crovisier (1978), or fixed  $R_0$ ,  $a_2$ , and  $a_3$  in the Reid et al. (2019) A5 model), yet these parameters have associated uncertainties that should be propagated into our inferred  $z$  distribution. A Bayesian model is well-suited to overcome these complications.

Here, we develop a Bayesian model, parameterized in terms of the  $z$  distribution moment ratio,  $\langle|z|^3\rangle/\langle|z|^2\rangle$ , to predict the LSR velocity distribution of H I absorbing clouds. As in the least-squares analyses, each cloud is assigned to its distance expectation value following Equation (13), at which we evaluate the expected LSR velocity from the Reid et al. (2019) A5 Galactic rotation model,  $V_{\text{model}}$ . By assigning each cloud to its distance expectation value, we are marginalizing

**Table 2**  
Least-squares Analysis Results

Sample	$N$	$U_{\odot}$ (km s $^{-1}$ )	$V_{\odot}$ (km s $^{-1}$ )	$W_{\odot}$ (km s $^{-1}$ )	$\ell_0$ (deg)	$\langle  z ^3 \rangle / \langle  z ^2 \rangle$ (pc)	$\sigma_V$ (km s $^{-1}$ )
Crovisier (1978) Reproduction; $ b  > 10^\circ$							
$ V_{\text{LSR}}  < 40 \text{ km s}^{-1}$	294	$-1.15 \pm 0.71$	$-0.20 \pm 0.83$	$-0.93 \pm 0.96$	$-10.95 \pm 2.96$	$132.74 \pm 14.07$	6.81
$ V_{\text{LSR}}  < 25 \text{ km s}^{-1}$	289	$-1.49 \pm 0.65$	$0.29 \pm 0.77$	$-0.93 \pm 0.88$	$-12.01 \pm 3.04$	$119.78 \pm 12.96$	6.23
$ V_{\text{LSR}}  < 15 \text{ km s}^{-1}$	271	$-1.18 \pm 0.53$	$0.69 \pm 0.63$	$-1.07 \pm 0.70$	$-15.09 \pm 3.57$	$85.56 \pm 10.73$	4.87
Crovisier (1978) Reproduction; $ V_{\text{LSR}}  < 25 \text{ km s}^{-1}$							
$10^\circ <  b  < 20^\circ$	119	$-0.33 \pm 1.07$	$0.70 \pm 1.21$	$-5.01 \pm 3.00$	$-11.24 \pm 4.06$	$111.72 \pm 16.45$	7.18
$20^\circ <  b  < 40^\circ$	102	$-2.16 \pm 0.95$	$-0.93 \pm 1.17$	$-1.37 \pm 1.30$	$-16.53 \pm 4.82$	$188.52 \pm 33.57$	5.10
Reid et al. (2019) Galactic Rotation Model; $ b  > 10^\circ$							
$ V_{\text{LSR}}  < 40 \text{ km s}^{-1}$	294	$11.65 \pm 0.73$	$15.88 \pm 0.85$	$8.60 \pm 0.98$	...	$128.04 \pm 14.88$	6.97
$ V_{\text{LSR}}  < 25 \text{ km s}^{-1}$	289	$11.97 \pm 0.67$	$15.38 \pm 0.79$	$8.61 \pm 0.90$	...	$114.25 \pm 13.82$	6.40
$ V_{\text{LSR}}  < 15 \text{ km s}^{-1}$	271	$11.61 \pm 0.54$	$14.90 \pm 0.65$	$8.76 \pm 0.72$	...	$77.85 \pm 11.62$	5.03
Reid et al. (2019) Galactic Rotation Model; $ V_{\text{LSR}}  < 25 \text{ km s}^{-1}$							
$10^\circ <  b  < 20^\circ$	119	$10.77 \pm 1.11$	$15.11 \pm 1.23$	$12.75 \pm 3.09$	...	$107.41 \pm 17.52$	7.40
$20^\circ <  b  < 40^\circ$	102	$13.15 \pm 0.99$	$16.28 \pm 1.24$	$8.45 \pm 1.37$	...	$166.91 \pm 39.89$	5.44

over the latent distance of each cloud, which we do not care to infer, under the assumption that it is poorly constrained by the velocity. We assume that the random component of the LSR velocity residuals is normally distributed with variance  $\sigma_V^2$ , so the likelihood that a cloud with observed LSR velocity  $V_i$  is drawn from the model is a normal distribution with mean  $V_i - V_{\text{model}}$  and variance  $\sigma_V^2$ . To account for the possibility of outliers in our data, we include a second, zero-centered, normally distributed component in the likelihood distribution with a variance  $\sigma_{V,\text{outlier}}^2$ . The total likelihood that cloud  $i$  is drawn from this mixture model is thus

$$P(V_{i,\text{LSR}} | \mathbf{w}, \boldsymbol{\theta}) = w_0 N(V_i - V_{\text{model}}, \sigma_V^2 | \boldsymbol{\theta}) + w_1 N(0, \sigma_{V,\text{outlier}}^2), \quad (17)$$

where  $\mathbf{w} = (w_0, w_1)$  are mixture weights that quantify the fraction of data belonging to the model or outlier population, respectively, such that  $w_0 + w_1 = 1$ , and  $\boldsymbol{\theta}$  is the set of model parameters that defines the model, including  $\langle |z|^3 \rangle / \langle |z|^2 \rangle$  and the Galactic rotation model parameters. We again ignore the LSR velocity uncertainties of individual clouds in this likelihood calculation, since they are typically much smaller than  $\sigma_V$  and thus do not contribute significantly to the likelihood determination.

In addition to the likelihood, we must also specify our prior knowledge of the model parameters. For the  $z$  distribution moment ratio, we adopt a  $k = 2$  gamma distribution with a scale parameter  $\theta = 50 \text{ pc}$ . The gamma distribution is a convenient specification both because it is broad (variance  $k\theta^2$ )—and thus encompasses a reasonable range of values—as well as because it approaches zero probability for unphysically small moment ratios. For the likelihood distribution widths,  $\sigma_V$  and  $\sigma_{V,\text{outlier}}$ , we use half-normal distributions with widths 10 and 50 km s $^{-1}$ , respectively, which again are quite broad and encompass physically reasonable expectations. The mixture weight prior is a Dirichlet distribution with equal concentration. Finally, for the Galactic rotation model parameters, we adopt as a prior a multivariate normal distribution with means equal to the Reid

et al. (2019) A5 model posterior point estimates and covariances inferred from the A5 model posterior distribution standard deviations and correlation matrix (M. Reid 2024, private communication). Random samples from these prior distributions (i.e., prior predictive checks) cover the range of observed data, and thus we are confident that the chosen prior distributions are reasonable.

We draw posterior samples from our Bayesian model using the Hamiltonian Monte Carlo No-U-Turn sampler (NUTS; Hoffman & Gelman 2014) implemented in *pyMC*, a probabilistic programming framework in *Python* (Oriol et al. 2023). To diagnose convergence and ensure that the sampler is not confined to local maxima of the posterior distribution, we draw 500 tuning samples followed by 500 posterior samples across four independent Markov chains. We check that the chains are converged by computing the  $\hat{R}$  statistic, which compares interchain variance to that across all chains (Vehtari et al. 2021), as well as the effective sample size of each parameter. In this and all subsequent MCMC analyses, we find  $\hat{R} < 1.01$  and effective sample sizes  $> 1000$  for all parameters, which is indicative of convergence. Furthermore, we draw random samples from the posterior distribution (i.e., posterior predictive checks) and confirm that the sampled posterior distribution is able to reproduce the observations.

The results of our moment ratio MCMC analysis are listed in the first column of Table 3. Here, we use all of the HI absorption spectra with  $|b| > 10^\circ$ ; we do not exclude the high-velocity features, since our model is capable of identifying them as outliers. The model parameters include the  $z$  distribution moment ratio,  $\langle |z|^3 \rangle / \langle |z|^2 \rangle$ , the fraction of data belonging to the outlier distribution,  $f_{\text{outlier}} \equiv w_1$ , the width of the model and outlier distributions,  $\sigma_V$  and  $\sigma_{V,\text{outlier}}$ , respectively, and the parameters that describe the Reid et al. (2019) A5 Galactic rotation model: the solar Galactocentric distance,  $R_0$ , the shape parameters,  $a_2$  and  $a_3$ , and the solar motion with respect to the LSR,  $U_{\odot}$ ,  $V_{\odot}$ , and  $W_{\odot}$ . Although we do not sample the shape parameters directly, we infer the Gaussian scale height,  $\sigma_z$ , exponential scale height,  $\lambda_z$ , and rectangular distribution half-width,  $W_{z,1/2}$ , from the moment ratio posterior distributions and include them in Table 3. A visual inspection

**Table 3**  
MCMC Analysis Results

Model	Moment Ratio	Gaussian	Exponential	Rectangular				
$N$	297	297	297	297				
$R_0$ (kpc)	8.225	$+0.144$ $-0.134$	8.220	$+0.146$ $-0.140$	8.228	$+0.139$ $-0.138$	8.222	$+0.158$ $-0.116$
$U_\odot$ (km s $^{-1}$ )	12.03	$+0.56$ $-0.48$	12.02	$+0.55$ $-0.52$	12.05	$+0.45$ $-0.63$	12.00	$+0.51$ $-0.58$
$V_\odot$ (km s $^{-1}$ )	14.40	$+0.76$ $-0.66$	14.36	$+0.76$ $-0.62$	14.37	$+0.68$ $-0.82$	14.37	$+0.63$ $-0.84$
$W_\odot$ (km s $^{-1}$ )	8.18	$+0.48$ $-0.56$	8.19	$+0.47$ $-0.56$	8.18	$+0.48$ $-0.59$	8.20	$+0.53$ $-0.52$
$a_2$	0.955	$+0.048$ $-0.044$	0.954	$+0.039$ $-0.053$	0.956	$+0.052$ $-0.040$	0.954	$+0.045$ $-0.043$
$a_3$	1.616	$+0.013$ $-0.014$	1.616	$+0.015$ $-0.012$	1.617	$+0.012$ $-0.015$	1.617	$+0.012$ $-0.014$
$\sigma_V$ (km s $^{-1}$ )	5.22	$+0.45$ $-0.38$	5.21	$+0.37$ $-0.46$	5.23	$+0.38$ $-0.45$	5.19	$+0.37$ $-0.44$
$f_{\text{outlier}}$	0.133	$+0.034$ $-0.048$	0.132	$+0.033$ $-0.051$	0.130	$+0.032$ $-0.050$	0.134	$+0.034$ $-0.047$
$\sigma_{V,\text{outlier}}$ (km s $^{-1}$ )	22.9	$+2.3$ $-4.9$	22.8	$+2.2$ $-4.9$	23.0	$+2.2$ $-5.1$	22.6	$+2.4$ $-4.8$
$\langle  z ^3 \rangle / \langle  z ^2 \rangle$ (pc)	97.0	$+14.9$ $-15.0$	98.4	$+15.4$ $-14.4$	100.0	$+13.5$ $-15.8$	95.7	$+12.4$ $-16.3$
$\sigma_z$ (pc)	60.8	$+9.3$ $-9.4$	61.7	$+9.6$ $-9.0$	...	...	...	...
$\lambda_z$ (pc)	32.3	$+5.0$ $-5.0$	...	33.3	$+4.5$ $-5.3$	...	...	...
$W_{z,1/2}$ (pc)	129.4	$+19.9$ $-20.0$	...	...	...	127.5	...	$+16.6$ $-21.7$
log ELPD	...	...	-1017.5	$\pm 20.3$	-1017.5	$\pm 20.2$	-1017.5	$\pm 20.3$

of the marginalized posterior distributions for each parameter reveals that they are unimodal and approximately normally distributed. The parameter values and uncertainties in Table 3 represent the posterior mean point estimates and 68% highest posterior density credible intervals, respectively.

### 3.5. MCMC Shape Parameter

If we assume the shape of the  $z$  distribution of H I absorbing clouds, then we can craft a Bayesian model to infer the shape parameter of that distribution directly. Such a model is beneficial in that it allows us to perform model comparison and determine which distribution shape best represents the data. The only difference between this model parameterization and the former is the relationship between the  $z$  distribution shape parameter and the distribution moment ratio. For a Gaussian, exponential, and rectangular distribution parameterized by Gaussian scale height  $\sigma_z$ , exponential scale height  $\lambda_z$ , and rectangular half-width  $W_{z,1/2}$ , respectively, the  $z$  distribution moment ratio,  $\langle |z|^3 \rangle / \langle |z|^2 \rangle$ , is  $2\sigma_z\sqrt{2/\pi}$ ,  $3\lambda_z$ , and  $3W_{z,1/2}/4$ , respectively. We use as a prior on the shape parameters a  $k=2$  gamma distribution with a scale parameter  $\theta=50$  pc, which again is a convenient distribution both because the prior probability goes to zero for unphysically small shape parameters as well as because the distribution is wide enough to cover reasonable shape parameter values. Prior predictive checks demonstrate that our choice of priors reasonably represents the data.

The results of our shape parameter MCMC analyses are listed in the last three columns of Table 3 for the Gaussian, exponential, and rectangular models. Again, we include all H I absorption features with  $|b| > 10^\circ$ . We list the derived  $z$  distribution moment ratio, although this parameter is not sampled directly for these models. All models converge after 500 tuning samples and 500 posterior samples as determined by  $\hat{R} < 1.01$  for each parameter, an effective sample size  $> 1000$  for each parameter, and a visual inspection of the posterior predictive checks. The marginalized posterior distributions are again unimodal and normally distributed in all cases, so the parameter values and uncertainties in Table 3 represent the posterior mean point estimates and 68% highest posterior density credible intervals, respectively.

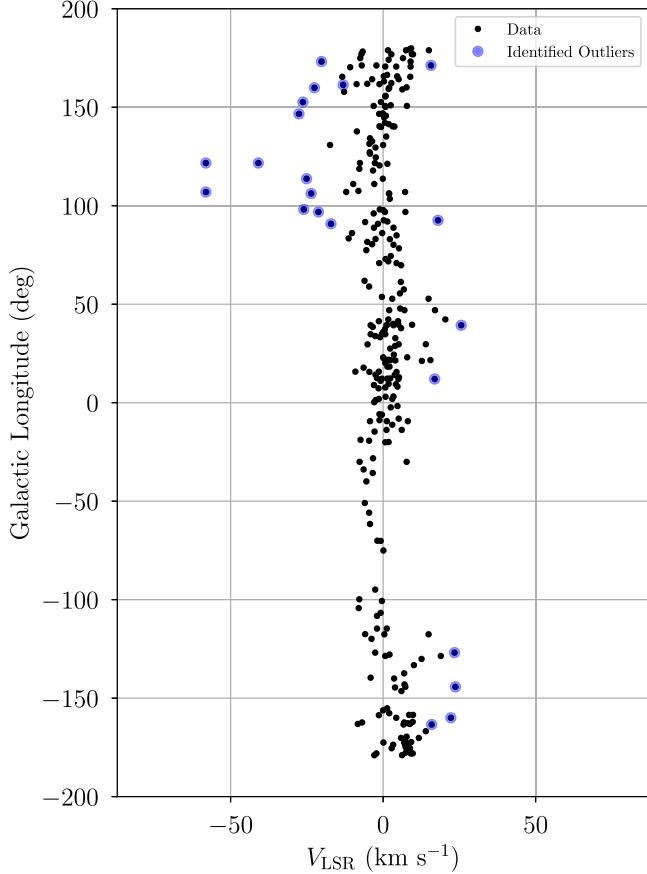
Finally, we determine which of the three assumed  $z$  distribution shapes best represents the data using leave-one-out cross-validation (Vehtari et al. 2017) as implemented in *ArviZ*, a *Python* package for exploratory analysis of Bayesian models (Kumar et al. 2019). For each shape parameter model, the last row of Table 3 lists the expected log pointwise predictive density (ELPD), which is a measure of the out-of-sample predictive fit. The model with the largest ELPD is preferred.

## 4. Discussion

### 4.1. Vertical Distribution

Although the statistical bias introduced by latitude truncation alters the interpretation of the Crovisier (1978) analysis, the statistical model that the author develops to infer the vertical distribution of H I absorbing clouds is robust. Based on our least-squares implementation of this model and using the original Crovisier et al. (1978) data, we are able to reproduce the Crovisier (1978) results and infer the shape of the  $z$  distribution of cold H I clouds. Depending on the LSR velocity selection of the subsample, we infer a  $z$  distribution shape defined by the ratio of the third moment (skewness) to the second moment (variance),  $\langle |z|^3 \rangle / \langle |z|^2 \rangle \simeq 110 \pm 15$  pc (see Table 2). For an assumed Gaussian distribution, this moment ratio corresponds to a scale height  $\sigma_z \simeq 70$  pc, which is a factor of two smaller than that inferred by Crovisier (1978), due to the latitude truncation bias discussed in Section 2.1.

The least-squares analysis results are sensitive to the LSR velocity criteria of each subsample, which implies that the data include outliers that are not explained by the model. Therefore, we develop two Bayesian models—one to infer the  $z$  distribution moment ratio and one to infer the shape parameter of an assumed distribution directly—that include an outlier component in the likelihood distribution. We find that each model converges to a similar  $z$  distribution moment ratio,  $\langle |z|^3 \rangle / \langle |z|^2 \rangle \simeq 97 \pm 15$  pc, which corresponds to a Gaussian scale height  $\sigma_z = 61 \pm 9$  pc, an exponential scale height  $\lambda_z = 32 \pm 5$  pc, and a rectangular half-width  $W_{z,1/2} = 129 \pm 20$  pc (see Table 3). The ELPD is nearly identical between models, which indicates that the data are not able to distinguish between these three assumed distributions. Furthermore, we find that  $\sim 13\%$  of the Crovisier et al. (1978) H I absorption detections appear to be poorly explained by our



**Figure 3.** Galactic longitudes and LSR velocities,  $V_{\text{LSR}}$ , of the Crovisier et al. (1978) H I absorption detections. Those data with a  $>50\%$  posterior probability of being outliers in the moment ratio model are highlighted in blue.

model and instead belong to an outlier population. Figure 3 shows the Galactic longitude–velocity distribution of the Crovisier et al. (1978) H I absorption detections, where we highlight those data with a  $>50\%$  probability of being a member of the outlier population based on samples from the posterior distribution of the moment ratio model. Clearly, the outliers belong to a population of intermediate-velocity H I absorption features. These detections may represent a collection of intermediate-velocity clouds (e.g., Wakker 2004; Lehner et al. 2022), perhaps the remnants of infalling high-velocity clouds (e.g., Begum et al. 2010), but we defer further investigation to a future work.

#### 4.2. Implications

The fact that we have reduced the apparent vertical distribution of cold H I clouds by a factor of two compared to the original Crovisier (1978) analysis has potentially significant implications for our understanding of hydrostatic balance in the solar neighborhood ISM. Our result suggests that, in the solar neighborhood, the scale height of the CNM is nearly the same as that of the molecular gas layer ( $\sigma_z \simeq 50$  pc, e.g., Heyer & Dame 2015). This complicates the already limited characterization of the vertical distribution of the CNM given the few observational constraints (McClure-Griffiths et al. 2023) and draws into question the claimed variation in CNM scale height with Galactocentric distance.

It is possible that the solar neighborhood gives a biased view of the vertical distribution of CNM clouds. Consider, for

example, the analysis of Dickey et al. (2022), which uses H I absorption detections near  $0 \text{ km s}^{-1}$  to infer the vertical distribution of cold H I clouds near the solar Galactocentric distance. They find a latitude distribution of H I absorption features composed of two Gaussian distributions, the broader of which they attribute to local H I and the narrower of which they attribute to H I at the solar Galactocentric distance on the far side of the Galaxy. Given a heliocentric distance  $d \simeq 15.6$  kpc, the latitude distribution width of  $\sim 0^{\circ}59$  corresponds to a  $z$  distribution scale height  $\sigma_z = 160$  pc. This is more than a factor of two greater than our inferred CNM scale height in the solar neighborhood. Since Dickey et al. (2022) select their sample kinematically, they may be biased by intermediate-velocity cloud interlopers that exist at a variety of distances and thus complicate the interpretation of the latitude distribution. It is also possible, however, that the CNM scale height varies due to local effects like the stellar surface density and star formation history.

Recent isolated galaxy simulations have attempted to overcome the resolution and physical limitations of galaxy evolution models in order to predict realistic distributions of different ISM phases on subparsec scales. One such simulation, that of Smith et al. (2023), uses the hydrodynamic galaxy models of Tress et al. (2020) and Tress et al. (2021) to reach parsec resolution (or better) in the cold gas. The Smith et al. (2023) simulation assumes a constant metallicity across the disk, but they test the effect of a varying FUV interstellar radiation field that is tied to the time-averaged star formation rate surface density distribution compared to a constant radiation field. In both cases, they find that the CNM scale height is smaller than the total H I scale height everywhere outside of the Galactic center. This is in contrast to the Dickey et al. (2009) study, which finds a similar scale height between the WNM and CNM in the outer Galaxy. Furthermore, the Smith et al. (2023) simulations demonstrate that, although the azimuthally averaged total H I scale height increases with Galactocentric distance, the azimuthally averaged CNM scale height remains roughly constant at  $\sigma_z \simeq 50$  pc for the varying radiation field and increases slightly from  $\sim 50$  to  $\sim 100$  pc for the constant radiation field. These simulations are thus consistent with our solar neighborhood inference, although their Figure 10 suggests that there could be local variations in the CNM scale height on the order of a factor of  $\sim 2$ . Such local variations could explain the Dickey et al. (2022) scale height at the solar Galactocentric distance in the fourth quadrant.

#### 4.3. Galactic Rotation

In addition to the vertical distribution of H I absorbing clouds, our Bayesian model also infers the Reid et al. (2019) Galactic rotation model parameters, including the solar Galactocentric distance. We use as priors a multivariate normal distribution defined by the means, standard deviations, and correlation coefficients of the Reid et al. (2019) A5 model posterior distribution. Except for the solar peculiar velocity components, our model posterior distribution means match the prior distribution means within the uncertainties, and the widths of the posterior distributions are similar to those of the priors (see Table 3). These two facts suggest that the Crovisier et al. (1978) data are not able to further constrain  $R_0$  and the Galactic rotation model parameters. On the other hand, both our least-squares analysis and Bayesian models converge to similar estimates for the solar peculiar velocity components,

$(U_\odot, V_\odot, W_\odot) = (12.0 \pm 0.5, 14.4 \pm 0.7, 8.2 \pm 0.5)$ , which are more precise than the prior distributions. These parameter estimates are consistent with those inferred from stellar kinematics (Schönrich et al. 2010; Zbinden & Saha 2019), although the stellar results appear sensitive to both the method of analysis and the stellar sample (e.g., Ding et al. 2019).

#### 4.4. Code Package

The Bayesian model, least-squares, and MCMC methods used this work are provided to the community as an open-source package: `kinematic_scaleheight`<sup>3</sup> (Wenger 2024b). Although we have only demonstrated it as applied to H I absorption data, it is possible to use these models and algorithms to infer the vertical distribution of any tracer in the solar neighborhood for which there is a reasonable expectation that the tracer follows a given Galactic rotation model.

### 5. Conclusion and Future Work

The vertical distribution of the ISM, perpendicular to the Galactic plane, is a constraint on models of ISM hydrostatic balance and galaxy evolution. In particular, the distribution of CNM clouds appears to be set gravitationally by the stellar surface density and kinematically by the energy injection from star formation (McKee & Ostriker 1977; Smith et al. 2023). Due to the observational challenges associated with identifying individual CNM clouds, few studies have explored their vertical distribution across the Galactic disk. In the solar neighborhood, the statistical analysis of Crovisier (1978) and the extension by Belfort & Crovisier (1984) are the only such studies.

We identify a bias in the Crovisier (1978) method that results in an overestimate of the scale height of local CNM clouds by a factor of  $\sim 1.5$  to  $\sim 3$ . With a corrected least-squares method as well as a novel Bayesian model, we use the Crovisier et al. (1978) H I absorption data to infer that the vertical distribution of H I absorption in the solar neighborhood is best described by  $\langle |z|^3 \rangle / \langle |z|^2 \rangle = 97 \pm 15$  pc, where  $\langle |z|^3 \rangle$  and  $\langle |z|^2 \rangle$  are the third (skewness) and second (variance) moments of the distribution, respectively. This moment ratio corresponds to a Gaussian scale height  $\sigma_z = 61 \pm 9$  pc, an exponential scale height  $\lambda_z = 32 \pm 5$  pc, and a rectangular half-width  $W_{z,1/2} = 129 \pm 20$  pc, although the Crovisier et al. (1978) data are unable to distinguish between these three possible vertical distribution shapes. Furthermore, our model suggests that  $\sim 13\%$  of the local H I absorption detections may belong to a population of intermediate-velocity cloud interlopers, since they are not well-described by our model. Both the least-squares and Bayesian models converge to estimates of the solar peculiar motion that are consistent with those inferred from stellar kinematics (e.g., Schönrich et al. 2010):  $(U_\odot, V_\odot, W_\odot) = (12.0 \pm 0.5, 14.4 \pm 0.7, 8.2 \pm 0.5)$ . Our analysis tools are provided to the community as an open-source software package (Wenger 2024b).

The scale height of CNM clouds in the solar neighborhood is similar to both the inferred inner-Galaxy CNM scale height ( $\sigma_z \simeq 50$ –90 pc, e.g., Dickey et al. 2022) as well as the scale height of the molecular gas layer in the solar neighborhood ( $\sigma_z \simeq 50$  pc, e.g., Heyer & Dame 2015). These results are consistent with simulations, which suggest that the scale height

of cold H I should remain roughly constant out to the edge of the star-forming disk (Smith et al. 2023). At the solar Galactocentric distance on the far side of the Galaxy, however, Dickey et al. (2022) find a  $2$ – $3 \times$  larger scale height. Either there are significant variations in the scale height due to local effects, or their inference is biased due to intermediate-velocity H I absorption interlopers that complicate the interpretation of their kinematically selected sample. A comprehensive analysis of modern H I absorption observations in the solar neighborhood is warranted and will be the subject of an upcoming study.

### Acknowledgments

This research has made use of NASA’s Astrophysics Data System Bibliographic Services. T.V.W. is supported by a National Science Foundation Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2202340. D.R.R. is supported by a National Science Foundation Astronomy and Astrophysics Postdoctoral Fellowship under award AST-2303902. The authors acknowledge Interstellar Institute’s program “II6” and the Paris-Saclay University’s Institute Pascal for hosting discussions that nourished the development of the ideas behind this work, as well as J. M. Dickey for providing comments that improved the quality of this paper.

*Software:* ArviZ (Kumar et al. 2019), pyMC (Oriol et al. 2023), Matplotlib (Hunter 2007), NumPy & SciPy (van der Walt et al. 2011), `kinematic_scaleheight` (Wenger 2024b).

### ORCID iDs

Trey V. Wenger  <https://orcid.org/0000-0003-0640-7787>  
 Daniel R. Rybacyk  <https://orcid.org/0000-0003-3351-6831>  
 Snežana Stanimirović  <https://orcid.org/0000-0002-3418-7817>

### References

Begum, A., Stanimirović, S., Peek, J. E., et al. 2010, *ApJ*, **722**, 395  
 Belfort, P., & Crovisier, J. 1984, *A&A*, **136**, 368  
 Bronfman, L., Cohen, R. S., Alvarez, H., May, J., & Thaddeus, P. 1988, *ApJ*, **324**, 248  
 Crovisier, J. 1978, *A&A*, **70**, 43  
 Crovisier, J., Kazes, I., & Aubry, D. 1978, *A&AS*, **32**, 205  
 Dickey, J. M., Dempsey, J. M., Pingel, N. M., et al. 2022, *ApJ*, **926**, 186  
 Dickey, J. M., Kulkarni, S. R., van Gorkom, J. H., & Heiles, C. E. 1983, *ApJS*, **53**, 591  
 Dickey, J. M., McClure-Griffiths, N., Gibson, S. J., et al. 2013, *PASA*, **30**, e003  
 Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, *ApJ*, **585**, 801  
 Dickey, J. M., Strasser, S., Gaensler, B. M., et al. 2009, *ApJ*, **693**, 1250  
 Ding, P.-J., Zhu, Z., & Liu, J.-C. 2019, *RAA*, **19**, 068  
 Field, G. B., Goldsmith, D. W., & Habing, H. J. 1969, *ApJL*, **155**, L149  
 Heyer, M., & Dame, T. M. 2015, *ARA&A*, **53**, 583  
 Hoffman, M. D., & Gelman, A. 2014, *Journal of Machine Learning Research*, **15**, 1593  
 Hopkins, P. F., Wetzel, A., Kereš, D., et al. 2018, *MNRAS*, **480**, 800  
 Hunter, J. D. 2007, *CSE*, **9**, 90  
 Kalberla, P. M. W., & Kerp, J. 2009, *ARA&A*, **47**, 27  
 Kavars, D. W., Dickey, J. M., McClure-Griffiths, N. M., Gaensler, B. M., & Green, A. J. 2003, *ApJ*, **598**, 1048  
 Kim, C.-G., Kim, W.-T., & Ostriker, E. C. 2010, *ApJ*, **720**, 1454  
 Kumar, R., Carroll, C., Hartikainen, A., & Martin, O. 2019, *JOSS*, **4**, 1143  
 Lehner, N., Howk, J. C., Marasco, A., & Fraternali, F. 2022, *MNRAS*, **513**, 3228  
 Levine, E. S., Blitz, L., & Heiles, C. 2006, *Sci*, **312**, 1773

<sup>3</sup> A copy of this model and our methods were deposited to Zenodo doi:10.5281/zenodo.10818724, Wenger (2024a).

McClure-Griffiths, N. M., Stanimirović, S., & Rybarczyk, D. R. 2023, *ARA&A*, **61**, 19

McKee, C. F., & Ostriker, E. C. 2007, *ARA&A*, **45**, 565

McKee, C. F., & Ostriker, J. P. 1977, *ApJ*, **218**, 148

Oriol, A.-P., Virgile, A., Colin, C., et al. 2023, *PeerJ Computer Science*, **9**, e1516

Ostriker, E. C., & Kim, C.-G. 2022, *ApJ*, **936**, 137

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2014, *ApJ*, **783**, 130

Reid, M. J., Menten, K. M., Brunthaler, A., et al. 2019, *ApJ*, **885**, 131

Schönrich, R., Binney, J., & Dehnen, W. 2010, *MNRAS*, **403**, 1829

Smith, R. J., Tress, R., Soler, J. D., et al. 2023, *MNRAS*, **524**, 873

Su, Y., Yang, J., Yan, Q.-Z., et al. 2021, *ApJ*, **910**, 131

Tress, R. G., Smith, R. J., Sormani, M. C., et al. 2020, *MNRAS*, **492**, 2973

Tress, R. G., Sormani, M. C., Smith, R. J., et al. 2021, *MNRAS*, **505**, 5438

van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *CSE*, **13**, 22

Vehtari, A., Gelman, A., & Gabry, J. 2017, *Statistics and Computing*, **27**, 1413

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. 2021, *BayAn*, **16**, 667

Wakker, B. P. 2004, in *High Velocity Clouds*, ed. H. van Woerden et al. (Dordrecht: Kluwer), 25

Wenger, T. V. 2024a, *tvwenger/kinematic\_scaleheight*: v1.0, Zenodo, doi:10.5281/zenodo.10818724

Wenger, T. V., 2024b *kinematic\_scaleheight*, Astrophysics Source Code Library, ascl:2403.003

Wolfire, M. G., McKee, C. F., Hollenbach, D., & Tielens, A. G. G. M. 2003, *ApJ*, **587**, 278

Zbinden, O., & Saha, P. 2019, *RNAAS*, **3**, 73