Bayesian inference of three-dimensional gas maps

I. Galactic CO

P. Mertsch and A. Vittino

Institute for Theoretical Physics and Cosmology (TTK), RWTH Aachen University, Sommerfeldstr. 16, 52074 Aachen, Germany
e-mail: pmertsch@physik.rwth-aachen.de

Received 6 April 2021 / Accepted 5 July 2021

ABSTRACT

Carbon monoxide (CO) is the best tracer of Galactic molecular hydrogen (H$_2$). Its lowest rotational emission lines are in the radio regime, and thanks to Galactic rotation, emission at different distances is Doppler shifted. For a given gas flow model, the observed spectra can thus be deprojected along the line of sight to infer the gas distribution. We used the CO-line survey of Dame et al. (2001, ApJ, 547, 792) to reconstruct the three-dimensional density of H$_2$. We considered the deprojection as a Bayesian variational inference problem. The posterior distribution of the gas densities allowed us to estimate the mean and uncertainty of the reconstructed density. Unlike most of the previous attempts, we took the correlations of gas on a variety of scales into account, which allowed us to correct for some of the well-known pathologies, such as finger-of-god effects. The two gas flow models that we adopted incorporate a Galactic bar that induces radial motions in the inner few kiloparsecs and thus offers spectral resolution towards the Galactic centre. We compared our gas maps with those of earlier studies and characterise their statistical properties, for instance the radial profile of the average surface mass density.

Key words. Galaxy: structure – ISM: kinematics and dynamics – ISM: molecules – methods: statistical

1. Introduction

Line surveys of molecular gas are a treasure trove for the study of the properties and dynamics of the interstellar medium (ISM; Ferrière 2001). While the dominant fraction of molecular gas is molecular hydrogen (H$_2$), its line emission is inefficient in the cold and dense phase because of the large spacing of its energy levels. The energy levels of carbon monoxide (CO), on the other hand, are much more closely spaced, rendering it radiatively more efficient in the same environments. To first approximation, the density of H$_2$ and CO are linearly related, making CO the preferred proxy for H$_2$. In particular the J = 1 → 0 emission of $^{12}$CO at 115 GHz has become the observable of choice in the study of the molecular ISM in the Galaxy. While $^{12}$CO is almost always in the optically thick regime, the emission from its isotopologue $^{13}$CO is in the optically thin regime, thus complementary (e.g. Szűcs et al. 2014), but still bright enough to allow mapping over large regions of the Galaxy (e.g. Schuller et al. 2021).

Ever since its observational discovery (Wilson et al. 1970), the coverage, sensitivity, and angular resolution of CO surveys have continuously improved. Such surveys have enabled the study of the molecular ISM on scales as small as individual star-forming clumps of a few solar masses up to the total Galactic H$_2$ mass of about 10$^9$M$_\odot$. The applications of CO surveys thus range widely. Studies of Galactic structure on the largest scales, in particular of features such as spiral arms or nuclear bars, are interesting in their own right (Hou & Han 2014), but also offer invaluable clues for the theory of galaxy evolution (Kewley et al. 2019). The processes at play in the formation and dissolution of molecular clouds can be investigated through the study of individual clouds, but also through the statistical analysis of cloud catalogues (e.g. Miville-Deschênes et al. 2017). As stellar nurseries, the regions of dense molecular gas are central in the study of star formation (Kennicutt & Evans 2012). Due to its superior angular resolution compared to broad-band emission in other wavelengths, the correlation between different emission processes relies on precise information on the molecular phases. One concrete example is diffuse emission produced by non-thermal cosmic rays (Ackermann et al. 2012). Here, the molecular gas (together with atomic gas) provides the target for high-energy cosmic rays, resulting in the production of non-thermal gamma-ray emission either through bremsstrahlung or through the production and subsequent decay of pions into high-energy gamma-rays and neutrinos. While the study of all such processes in the Galaxy is interesting in its own right, it is also important to study and calibrate the correlations with other processes in order to use the observations of CO lines for extragalactic astrophysics (e.g. Sun et al. 2018).

As a result of our vantage point in the Galaxy, the galactic distribution of CO and H$_2$ is not readily available from the gas-line surveys. However, because of Galactic rotation, different emission points along a line of sight in general posses different relative velocities with respect to the observer. Thus, the emission along a line of sight consists of a spectrum that encodes the distribution of emission with distance. The data products of gas-line surveys thus consist of spectra for individual lines of sight, oftentimes provided on a three-dimensional grid in longitude $\ell$, latitude $b$, and velocity with respect to the local standard of rest (LSR), v$_{LSR}$. For a given Galactic rotation curve, or more generally, given a gas flow model, such spectra can be deprojected...
in principle to find the three-dimensional distribution of CO and H$_2$.

Unfortunately, several complications hamper the deprojection of the 3D-cubes of gas-line surveys into a three-dimensional $xyz$-cube of gas densities. First, along a given line of sight and for a given velocity, most gas-flow models exhibit two distance solutions inside the solar circle. For the velocity range affected by this effect, it is a priori unclear what fractions of the emission reside at the near and far distances. This effect is commonly referred to as the kinematic distance ambiguity. Second, for circular rotation, sight lines close to the Galactic centre (longitude $\ell \approx 0^\circ$) and anti-centre (longitude $\ell \approx 180^\circ$) directions exhibit little to no radial velocity, thus lacking kinematical resolution: All the emission piles up around $V_{LSR} = 0$ km s$^{-1}$ and cannot be deprojected along the line of sight. Finally, peculiar velocities, that is, smaller-scale motions of gas on top of the large-scale gas flow, for instance due to stellar winds, supernova explosions, or spiral structures, perturb the smooth mapping of distance to radial velocity. These perturbations become visible as artefacts in the deprojected gas maps. Oftentimes, the inferred distribution of gas becomes smeared out along the line of sight, leading to the famous finger-of-god effect.

We illustrate the first two of these issues for the simple (and unrealistic) example of purely circular rotation. In this case, the radial velocity $v_{LSR}$ along $(\ell, b)$ for emission from the galactocentric radius $R$ is given by

$$v_{LSR}(R, \ell, b) = \cos b \sin \ell \left( \frac{R}{R_\odot} V(R) - V_\odot \right).$$  

(1)

Further assuming a flat rotation curve, $V(R) = 220$ km s$^{-1}$, we show $v_{LSR}$ as a function of position in the Galactic plane ($b = 0^\circ$) in Fig. 1. We have assumed the observer to be located at Cartesian coordinates $(x, y) = (R_\odot, 0)$ with $R_\odot = 8.15$ kpc, and the relation from longitude $\ell$, latitude $b$, and distance along the line of sight $s$ to the Cartesian coordinates $x$, $y$, and $z$ is

$$x = s \cos \ell \cos b - R_\odot,$$

(2)

$$y = s \sin \ell \cos b,$$

(3)

$$z = s \sin b.$$  

(4)

The lack of kinematic resolution around the Galactic centre and anti-centre direction is readily visible: each point along $\ell \approx 0, \pm 180^\circ$ is observed without any Doppler shift because there is no velocity component along the line of sight. It is also evident that the near-far ambiguity only affects positions within the solar circle, the circle centred on the Galactic centre with radius $R_\odot$. Here, a given velocity $v_{LSR} < 0$ ($v_{LSR} > 0$) corresponds to two solutions for $-90^\circ < \ell < 0^\circ$ ($0^\circ < \ell < 90^\circ$). The two solutions are equidistant from the intersection of the line of sight with the tangent point circle: a circle of radius $R_\odot/2$, centred at $(x, y) = (R_\odot/2, 0)$, the locus of points that are tangent to lines of constant longitude. For a given line of sight, the tangent point is an extremum of $v_{LSR}$.

Several previous attempts have been made to deproject the results of gas-line surveys into three-dimensional gas distributions, and each study had to adopt a way for solving the issues discussed above. As for the lack of kinematic resolution towards the Galactic centre and anti-centre, Nakanishi & Sofue (2006) adopted a circular gas-flow model, thus ruling out the possibility of directly reconstructing gas near $\ell = 0^\circ$ and $\ell = 180^\circ$. Instead, they resorted to interpolations between sidelines with low but finite radial velocities, at least for this side of the Galactic centre. The regions beyond the Galactic centre with little to no kinematical resolution were excluded. Pohl et al. (2008) instead adopted the result of a numerical simulation (Bissantz et al. 2003) of the gas flow that includes non-circular motions. Towards the Galactic centre and anti-centre, this provides several finite kinematic solutions, which even aggravates the distance ambiguity, but also provides some kinematic resolution. Gas could thus also be constructed close to $\ell = 0^\circ$ and $\ell = \pm 180^\circ$ without interpolation.

As for the near-far ambiguity, when we assume an exponential or Gaussian distribution of gas in the direction perpendicular to the plane of the disk, the emission at a fixed velocity in the inner Galaxy will have a latitude profile that is composed of two Gaussians. Performing fits to the latitude distribution in individual velocity bins thus allows us to determine the relative distribution of gas between the two distance solutions. This method is know as the double-Gaussian method, and a number of studies have adopted this technique (e.g. Clemens et al. 1988). Both Nakanishi & Sofue (2006) and Pohl et al. (2008) have also used the double-Gaussian method to break the near-far degeneracy in their deprojection. Specifically, Pohl et al. (2008) iteratively deprojected limited velocity ranges, assuming a certain thermal width, until the residual velocity spectra agreed with observational noise. Some artefacts are clearly still present in their gas maps, however. Nakanishi & Sofue (2006), on the other hand, were able to suppress some potential artefacts, but at the cost of adopting a rather coarse resolution, given the level of detail and small-scale structure available in the CfA CO survey compilation.

An alternative to the deprojection that is able to evade the problems described above to a certain degree is forward-modelling. For this, a parametric model for the gas distribution needs to be produced. Recently, Jöhnnesson et al. (2018) provided such models for atomic and molecular hydrogen and determined the free parameters by fits to existing survey data. The near-far ambiguity and the lack of kinematic resolution are both fixed in this case by assuming a coherent distribution of gas throughout the offending regions. The authors were able to constrain a large number of parameters by their fit, and the properties of the spiral arms bear some resemblance due to other complementary data sets. However, as the authors themselves admit, not all the gas implied from the survey data can be successfully
deprojected, such that the gas maps are to be considered as a lower limit of the real gas maps.

An important additional physical constraint that the gas densities have to fulfil but which is not leveraged by any of the previous studies is the existence of correlations. These correlations must exist on a range of scales and are due to a variety of processes. On the largest scales, they are due to the large-scale structure of the disk and the spiral arms, thus they ultimately are consequences of the formation history and density waves (Shu 2016). On smaller scales, the correlations are determined by the turbulent nature of the interstellar medium (Kolmogorov 1941), affecting both the fluctuation of gas density and velocity, and also magnetic fields. When the three-dimensional gas density is modelled as a (Gaussian) random field, these correlations can be parametrised by the power spectrum of gas densities. There is some support for the hypothesis of a log-normal random field (Nordlund & Padoan 1999; Ostriker et al. 2001), thus we considered the log of gas densities to behave like a Gaussian random field. This reconstruction problem for the gas density under the priors of a given correlation structure is most conveniently formulated in a Bayesian framework. We note that with a general enough inference method, the parameters for the power spectrum do not need to be assumed, but can be determined by the inference method together with the gas density. In addition, the Bayesian inference method provides not only an estimate of the gas density, but also quantifies its uncertainty. To our knowledge, none of the previous studies provided such an uncertainty estimate.

When the existence of correlations is taken into account, it is hoped that gas densities for regions for which data are less constraining may be predicted, for example, along \( \ell = 0^\circ \) and \( \ell = \pm 180^\circ \). The quality of the deprojected data for these regions is reflected in an increased uncertainty for these regions. Additionally, it can help break the near-far ambiguity: Because of the coherence on large scales, the two solutions with gas at different distances will not exhibit the same likelihood, and the inference algorithm will thus be able to distinguish between these solutions. Finally, taking the presence of observational noise into account in our inference method naturally allows denoising the observations.

The remainder of this paper is organised as follows: in Sect. 2 we briefly review the survey data, introduce the two gas-flow models we adopt, and present the Bayesian method adopted for the deprojection and our data model. Our results are shown in Sect. 3 in a variety of representations. We compare our gas maps and supplemental results with those of previous studies. We conclude in Sect. 4 and provide some thoughts about future directions of research. Details and supplemental information for one of the gas-flow models are provided in Appendix A.

2. Method

2.1. Survey data

We used the \(^{12}\text{CO} \ (1 \rightarrow 0)\) line spectra as compiled by Dame et al. (2001). They comprise 37 individual surveys that were performed with two 1.2 meter millimeter-wave telescopes operated at Columbia University in New York City, NY, at the Centre for Astrophysics in Cambridge, MA, and at Cerro Tololo in Chile. Together, these surveys cover the entire Galactic plane in longitude, extending to \( \pm 30^\circ \) in latitude, thus covering virtually all areas for which significant emission has been reported. The velocity range is from \(-319.8\) to \(319.8\) km s\(^{-1}\) in velocity. Between individual surveys, the angular resolution varies from 1/16 to 1/2, as does the rms noise level, which we conservatively fixed to 0.3 K per channel. The survey data were downloaded from the SAO Radio Telescope Data Centre\(^1\).

Before publication, the raw survey data had been corrected for the motion of the Earth and the Sun with respect to the LSR. Since the time of publication of Dame et al. (2001), better estimates of these relative velocities have become available, however. Therefore we corrected the survey data by taking the updated parameter values into account as described in Sect. 4.2 of Wenger et al. (2018), adopting the parameter values of a recent parallax-based determination of the distances to \( \sim 200 \) masers (Reid et al. 2019). Specifically, we converted the standard Cartesian velocity component \((U, V, W) = (10, 15, 7)\) km s\(^{-1}\) of the original surveys with the values \((\tilde{U}, \tilde{V}, \tilde{W}) = (10.6, 10.7, 7.6)\) km s\(^{-1}\) of model A5 of Reid et al. (2019). In Fig. 2 we show two projections of the corrected \((b, \ell)\)-cube that is integrated over velocity, the so-called zero-moment map (top panel), and integrated over latitude, the \( \ell - v \) diagram (bottom panel). In particular the \( \ell - v \) diagram nicely illustrates the coherent structures that will help remedy some of the deficiencies of the usual deprojection techniques.

After the deprojection, we converted from the inferred CO emissivity into \( \text{H}_2 \) gas density, adopting a linear relation between them. This relation is commonly assumed to exist between the \( \text{H}_2 \) column density \( N_{\text{H}_2} \) and the velocity-integrated CO brightness temperature \( T_{\text{CO}} \equiv \int \text{d}v T_b \), \( N_{\text{H}_2} = X_{\text{CO}} T_{\text{CO}} \), thus establishing the conversion factor \( X_{\text{CO}} \). While the relative abundances of \( \text{H}_2 \) and CO depend on a number of local factors (e.g. density, temperature, and metallicity), it is customary to adopt an average value of about \( 10^{20} \) molecules cm\(^{-2}\) (K km s\(^{-1}\))\(^{-1}\). The value \( X_{\text{CO}} = 2 \times 10^{20} \) molecules cm\(^{-2}\) (K km s\(^{-1}\))\(^{-1}\) with an uncertainty of \( \pm 30\% \) has been recommended (Bolatto et al. 2013). For a recent review of molecular gas surveys, see Heyer & Dame (2015).

2.2. Gas flow models

In addition to the three-dimensional gas density field that we wish to reconstruct, the distribution of CO brightness temperature in \( \ell, b, \) and \( v \) depends on the three-dimensional velocity field, which is also unknown. While it might seem that the assumption of purely circular rotation is the most model-independent assumption possible, it is problematic in two ways: practically, as it does not provide any kinematic resolution towards the Galactic centre and anti-centre directions, and factually, as the gas flow is known not to be purely circular in the inner Galaxy because of the well-established presence of the Galactic bar (Blitz & Spergel 1991) that induces radial motions in the inner few kiloparsec. While the existence of non-circular motions is qualitatively not challenged, the details are everything but certain.

In particular, we caution that the use of velocity fields based on simulations can be the source of significant artefacts if the modelled gas velocity differs from the true one. While this might seem an obvious point, it is easy to underestimate its importance because the effect of incorrect gas velocities on the reconstructed gas maps is aggravated by the particular form of the mapping between velocity \( v_{\text{LSR}} \) and distance \( x \) for regions with large gradients \( \partial v_{\text{LSR}} / \partial s \), gas observed over velocity intervals \( \Delta v_{\text{LSR}} \) is reconstructed in relatively narrow intervals of the line-of-sight distance \( \Delta s \), leading to regions of increased density. This can

\(^{1}\) https://www.cfa.harvard.edu/rtdc/
be considered the complement of an effect known as “velocity crowding”. Simulated spiral arms typically exhibit these velocity gradients, and their positions will therefore be accentuated by increased gas densities. Even though the simulations that we adopted as a basis for our gas-flow models below were adjusted in order to reproduce the results from gas-line surveys to some extent, we currently cannot prevent these artefacts. In order to estimate the systematic uncertainty they cause, we therefore adopted two gas-flow models.

The first model is the result of a smoothed particle hydrodynamics simulation (Bissantz et al. 2003, hereafter BEG03). In addition to the gravitational potential of bulge and bar, some assumptions were made about the presence of spiral arms, in particular, a four-armed spiral structure was assumed. The Sun was assumed to be at a distance of 8 kpc from the Galactic centre, and an angle of 20° was adopted for the inclination of the major axis of the bar and the line Sun-Galactic centre. The resulting $\ell$-v diagrams are similar to the CO data, but clearly miss some of the finer structure of the Dame et al. (2001) composite survey. We show their distribution of radial velocities in the top panel of Fig. 3. We extended the gas-flow model with a flat rotation curve beyond 8 kpc, in a similar fashion as Pohl et al. (2008). The assumed spiral arms are clearly visible in gradients in the velocity field. While newer simulations are available (e.g. Baba et al. 2010; Pettitt et al. 2014), we focussed here on BEG03 to allow a comparison of our results with those by Pohl et al. (2008), who used the same gas-flow model.

In addition, we constructed a gas-flow model based on a semi-analytical model for gas-carrying orbits in the potential dominated by the Galactic bar (Sormani et al. 2015, hereafter SBM15). We provide details of our model in Appendix A. The observer is assumed to be at a distance of 8.15 kpc from the Galactic centre (Reid et al. 2019), and the inclination of the major axis of the bar with respect to the line Sun-Galactic centre is again 20°. We show the resulting radial velocity map in the middle panel of Fig. 3 after rescaling the radial coordinate so as to bring the observer to the same galactocentric radius as in the BEG03 model in order to facilitate comparison. We also show the difference between the velocity fields of BEG03 and

---

Fig. 2. Two-dimensional projections of the data from the Dame et al. (2001) compilation of surveys, after correcting for the updated local parameters. Top: velocity-integrated skymap of Galactic CO emission. Bottom: $\ell$-v diagram of Galactic CO emission, integrated over latitudes from $-30^\circ$ to $30^\circ$. 

---

$^2$ Velocity crowding describes the effect that a large distance interval $\Delta s$ maps to a small velocity interval $\Delta v_{\text{LSR}}$ and vice versa for regions along the line of sight in which the radial velocity gradient $dv_{\text{LSR}}/ds$ is small.
2.3. Bayesian inference

In deprojecting gas line surveys, we followed two connected goals: First, we wish to reconstruct the three-dimensional gas density under the constraint that the density possesses a certain spatial correlation structure. While we can be sure that these correlations exist, the details are poorly constrained, and ideally, we would hope for the data to constrain the correlation structure itself. Second, we would also like to obtain an estimate of the uncertainties. These two goals are most directly achieved by adopting a Bayesian framework.

We considered the deprojection of the three-dimensional gas density from line-survey data as a high-dimensional Bayesian inference problem, that is, we sought the posterior distribution of the gas density for the given the survey data. According to Bayes’ theorem (Bayes & Price 1763), the posterior is proportional to the likelihood, that is, the probability for the observed brightness temperature given the gas density, times the prior, that is, the probability of the gas density. We could constrain ourselves to finding the maximum of the posterior, but the maximum can be uninteresting if the posterior is multi-modal or has degenerate directions. In addition, we would need to estimate the uncertainty separately, for instance, by adopting the Laplace approximation (Laplace 1866), which equates the covariance at the maximum posterior position with the inverse Hessian. Instead, it is advantageous to keep track of the uncertainty while exploring the posterior.

Monte Carlo Markov chain (MCMC) methods (Hastings 1970) do exactly this, and they can approximate arbitrary posteriors given large enough sample sizes. However, with growing dimensionality, they become computationally expensive, and for the large dimensionality of the current problem, prohibitively so. Variational inference (Blei et al. 2017) instead approximates the posterior with a parametric distribution, for instance a multivariate Gaussian. The parameters of the parametric distribution can be determined if the “distance” between the approximate distribution and the true posterior can be estimated, for instance through the Kullback-Leibler divergence (Kullback 1968). For a multi-variate Gaussian, this would in principle involve the inversion of the large covariance matrix, which is again computationally prohibitive. Instead, it has been suggested (Knollmüller & Enßlin 2019) to approximate the covariance with the inverse Fisher information metric, a method known as metric Gaussian variational inference. This method has recently been applied to problems ranging from reconstruction of the three-dimensional dust density in the Galaxy from reddening data (Leike & Enßlin 2019) to radio interferometry (Arras et al. 2019).

In practice, this is implemented as an iterative scheme, alternating between estimating the covariance at the current mean and updating the mean for the current estimate of the covariance. Specifically, adopting standardisation of the parameters, the computation of the (inverse) Fisher information metric requires the SBM15 model in the bottom panel of Fig. 3. Three major differences are evident: First, while the overall agreement inside the solar circle is good, the agreement in the inner 2 kpc is less so. Second, the perturbations due to the spiral structure present in the model of BEG03 are clearly absent in our SBM15 model. The difference is mostly of about ±10 km s⁻¹, but can be as large as ±30 km s⁻¹ in limited regions. Finally, outside the solar circle, the difference in the adopted rotation curves is marked and can lead to differences as large as ±30 km s⁻¹. We note, however, that there is little molecular gas at the relevant distances.

Fig. 3. Comparison of the two gas flow models that we adopted. Top: \(v_{\mathrm{LSR}}\) for the BEG03 model. Middle: \(v_{\mathrm{LSR}}\) for the SBM15 model. Bottom: difference between \(v_{\mathrm{LSR}}\) in the BEG03 and the SBM15 model. In all panels, the position of the Sun is marked with a cross. The color scale is different in the bottom panel.

\[\begin{align*}
\text{Top: } & v_{\text{LSR}} \text{ for the BEG03 model.} \\
\text{Middle: } & v_{\text{LSR}} \text{ for the SBM15 model.} \\
\text{Bottom: } & \text{difference between } v_{\text{LSR}} \text{ in the BEG03 and the SBM15 model. In all panels, the position of the Sun is marked with a cross. The color scale is different in the bottom panel.}
\end{align*}\]
we also show lines of constant longitude for the observer at \((x, y) = (R_0, 0)\).

We now illustrate two ways of how the assumed correlation structure helps with reconstructing individual features in the gas density. Consider the dark green contours around \((x, y) = (6, 1.8)\) kpc, which are the highest simulated density feature in this part of the plane. This structure is clearly located at the near distance. In a gas-line survey, this complex would be detected around longitude \(\ell \approx -40^\circ\) and at velocities \(v_{\text{LSR}}\) around \(-50\) and \(-40\) km s\(^{-1}\). Because of the kinematic distance ambiguity, however, emission around \(\ell \approx -40^\circ\) and \(v_{\text{LSR}} = \pm 50\) and \(-40\) km s\(^{-1}\) could also be due to a structure at the far distance, in this case, roughly centred around \((x, y) = (0, 6)\) kpc. We illustrate a structure that would contribute to line-survey data in a similar way by the red lines in Fig. 4. This structure is not part of the simulated density. It is very clear, however, that this structure is significantly distorted with respect to the original structure, in particular, it is much larger and very anisotropic. Such a large structure typically does not exist in the simulated density given the power spectrum that we adopted and will also not exist for the power spectrum that is inferred from the rest of the data in other parts of the plane. In the Bayesian inference scheme, the reconstruction of the structure at the far distance would therefore be penalised by the prior.

Second, the peak at the near distance, shown in dark green contours in Fig. 4, forms part of a larger extended structure, shown in light green contours. This is typical of the correlations at kiloparsec distances as parametrised by the power spectrum. Importantly, the kinematic distance of this extended structure is less uncertain because parts of this structure are close to the tangent point circle, where the kinematic distance is unambiguous. We note that the red contours, that is, the hypothetical reconstruction at the far distance, do not form part of this extended structure. (To see this, compare the red contours with the grey contours in its neighbourhood.) The prior therefore further penalises the reconstruction of this structure at the far distance.

The situation would be completely analogous for a structure that was truly at the far distance, in the sense that the typical size of the feature and its embedding in larger complexes would then lead to a preference of reconstruction at the far distance. If there is no clear preference of the near or far distance from the correlation structure, the inferred gas density maps will reflect this by an increased uncertainty at the respective near and far distance positions.

### 2.3.2. Data model

The relation between the CO emissivity \(\varepsilon(x, y, z)\) that we wish to reconstruct and the brightness temperature \(T(\ell, b, v)\) from the Dame et al. (2001) survey is given by a linear map \(R\) from signal space \((x, y, z)\) to data space \((\ell, b, v)\),

\[ R[\varepsilon](\ell, b, v) = \int_0^\infty ds \varepsilon(r) \delta(v - v_{\text{LSR}}(r)) \bigg|_{r = r(\ell, b, s)}, \tag{5} \]

with \(r(\ell, b, s)\) specified in Eqs. (2)–(4) above. Setting \(T = R[\varepsilon]\) would result in a deterministic likelihood \(p(T|\varepsilon)\),

\[ p(T_{\text{obs}}|F_{\text{sky}}) = \delta(T_{\text{obs}} - R[\varepsilon_{\text{sky}}]), \tag{6} \]

but we have to take additive noise \(n\) into account, which alters data model to

\[ T_{\text{obs}} = R[\varepsilon_{\text{sky}}] + n_{\text{obs}}. \tag{7} \]
We assumed this noise to be Gaussian distributed, \( n \sim p(n) = \mathcal{N}(n, N) \), with covariance \( N \) that is diagonal in harmonic space, \( \hat{N} = 2\pi \sigma_n^2 \delta(k - k') \), that is, white noise. Given the properties of the individual surveys combined in Dame et al. (2001), we fixed \( \sigma_n = 0.3 \) K. We marginalised over the noise, thus obtaining a Gaussian likelihood,

\[
\hat{p}(T_{\text{obs}} | \mathcal{G}) = \int dn p(T_{\text{obs}} | \mathcal{G}, n) p(n) \propto \int dn \delta \left( T_{\text{obs}} - R[\mathcal{G}] - n \right) \mathcal{G}(n, N) \tag{8}
\]

Finally, we took into account that the measured \( v \) differs from the actual radial velocity \( \dot{v} \) of the emitting gas, for example due to thermal line width or turbulence. We took the difference \( v - \dot{v} \) to be normal distributed with variance \( \sigma_v^2 \) and fixed \( \sigma_v = 5 \) \( \text{km s}^{-1} \). In principle, this requires another marginalisation, but we approximated this through a smearing of the linear map \( R \) instead,

\[
p(T_{\text{obs}} | \mathcal{G}) \equiv \int d\dot{v} \int dn \delta \left( T_{\text{obs}} - R[\mathcal{G}] - n \right) \mathcal{G}(n, N) \equiv \int dn \delta \left( T_{\text{obs}} - R[\mathcal{G}] - n \right) \mathcal{G}(n, N) \tag{10}
\]

We stress that this representation of \( P(k) \) is flexible enough to closely approximate the true underlying power spectrum for \( k \) where the data are constraining enough; where they are not, the shape is interpolated or extrapolated.

Finally, assuming \( s \) to be a homogeneous random field must fail when the deviations from this assumption for the real gas density become too strong. This is certainly the case when considering the confinement of molecular gas to the Galactic plane. In order to not be biased in the reconstruction, we scaled the log-normal signal field \( \mathcal{G} \) with an exponential profile in \( z \), \( \exp[-|z|/z_0] \). We adjusted \( z_0 \) to values such that the signal field \( \mathcal{G} \) averaged over different portions of the Galactic disk does not show strong gradients in the \( z \) direction and found \( z_0 = 40 \) pc to give satisfactory results. We also tested a Gaussian profile \( \exp[-z^2/(2\sigma_z^2)] \), but found the results to be virtually unchanged.

### 2.3.3. Details of the implementation

Because we performed our computations on a computer, the signal field is not a continuous field of gas densities, but rather a discretised version thereof. We adopted a Cartesian grid with the \( x, y, \) and \( z \) coordinates indexed with \( \alpha, \beta, \gamma \), that is, \( \mathcal{G}_{\alpha \beta \gamma} = \mathcal{G}(x_{\alpha}, y_{\beta}, z_{\gamma}) \). Specifically, we considered a \( 512 \times 512 \times 16 \) Cartesian grid stretching from \(-16 \) kpc to \( 16 \) kpc in the \( x \) and \( y \) directions and from \(-0.5 \) kpc to \( 0.5 \) kpc in the \( z \) direction, thus achieving a spatial resolution of \( 1/16 \) kpc = \( 62.5 \) pc.

Because the data are the spectra from a binned survey, they are already discretised, that is, \( T_{ij} = T(t_i, b_j, v_k) \). We degraded the data from their native resolution on a \( 2880 \times 481 \times 493 \) grid to a \( 1440 \times 241 \times 247 \) grid. The discrete version of the linear map of Eq. (5) is

\[
R_{\alpha \beta \gamma}^{\text{obs}} \equiv \int d\dot{v} \int dv \int ds \mathcal{G}(v - \dot{v}, \sigma_v^2) R[\mathcal{G}] \equiv \mathcal{G}(v - \dot{v}, \sigma_v^2)_{\text{r=r}_{\text{obs}}} \tag{18}
\]

To be able to define the posterior distribution, we still need to specify the signal prior. We modelled the CO emissivity \( \varepsilon(r) \) as a log-normal distributed random field, that is, \( s(r) \equiv \ln(\varepsilon(r)/\varepsilon_0) \) is normal distributed. Under the assumption that \( s \) is statistically homogeneous, that is, the two-point correlation in configuration space, \( \langle s(r)s(r') \rangle \), is a function of the distance \( |r - r'| \) only, the two-point correlation in the harmonic domain becomes “diagonal”, that is, \( \langle \mathcal{G}(k)\mathcal{G}(k') \rangle = 2\pi P(k) \delta(k - k') \), where \( P(k) \) is the power spectrum. We further assumed that the power spectrum \( P(k) \) is isotropic, \( P(k) = P(k) \).

Instead of assuming a form for the power spectrum \( P(k) \), we would like to determine it during the reconstruction as well. Following Leike & Enßlin (2019), we therefore adopted a statistical model for the power spectrum, with a Gaussian-distributed normal \( q \), a Gaussian distributed power-law index \( m \), and add a Gaussian random field in log(k),

\[
\sqrt{P(k)} = \exp \left( \mu_\alpha + \sigma_\alpha \phi_\alpha \right) + \left( \mu_m + \sigma_m \phi_m \right) \log(k) + \mathcal{F}^{-1} \left[ A \left( 1 + T^2/2\Delta \right) \tau(t) \right] \tag{16}
\]

\[
A64, page 7 of 14
\]

3. Results and discussion

The main result of our analysis are the three-dimensional gas maps obtained from the two gas-flow models and the corresponding uncertainties. Our gas maps provide the highest-resolution three-dimensional deprojections of CO gas-line surveys to date, with a number of robust and well-localised emission regions. They should prove useful in the study of Galactic structure and diffuse emission. We make our maps available to the community.

In Fig. 5 we show the projection of the \( \text{H}_2 \) density onto the Galactic plane and its standard deviation for the BEG2003 gas-flow model (top panels) and the SMB15 gas-flow model (bottom panels). For each gas-flow model, the survey data were successfully deprojected into localised clusters of emission. The total reconstructed gas mass is \( 1.4 \times 10^9 \) \( M_\odot \) for the BEG03 model and \( 1.3 \times 10^9 \) \( M_\odot \) for the SMB15 model. This largely agrees with the value of \( (1 \pm 0.3) \times 10^9 \) \( M_\odot \) quoted by Heyer & Dame (2015).

Some elongated structures, spurs or spiral arm segments, are immediately visible. Some of the structures are

\[
3 \text{ https://gitlab.mpcdf.mpg.de/ift/nifty}
\]

\[
4 \text{ http://dx.doi.org/10.5281/zenodo.4485437}
\]
Fig. 5. Two-dimensional projection of reconstructed three-dimensional maps of molecular hydrogen. Top left: mean gas surface density $\Sigma$ for the BEG03 gas-flow model. Top right: standard deviation of the gas surface density $\Sigma$ for the BEG03 gas-flow model. Bottom left: mean gas surface density $\Sigma$ for the SBM15 model. Bottom right: standard deviation of the gas surface density $\Sigma$ for the SBM15 model.

more easily visible for the BEG03 model than for the SBM15 model. For instance, the two vertical spurs stretching from $(x, y) = (4, -1)$ kpc to $(x, y) = (4, 3)$ kpc and from $(x, y) = (6, -1)$ kpc to $(x, y) = (6, 3)$ kpc, respectively, are easily identified for the BEG03 model (top left panel of Fig. 5), but blend into a more extended emission for the SBM15 model. Revisiting Fig. 3 and in particular its bottom panel, it is clear that these structures are oftentimes linked to local extrema in the radial velocity field. For instance, the two spurs discussed above coincide with maxima of the velocity field of the BEG03 model, but are absent in the SBM15 model. They are due to the spiral arms, which have been put in by hand in the case of the BEG03 model, but not in the SBM15 model. We can also identify a spur that coincides with the tangent point circle, stretching from $(x, y) = (2, -3)$ kpc to $(x, y) = (6, -4)$ kpc. This is likely an artefact that indicates that the velocity range in the SBM15 model is too narrow.

Furthermore, a word of caution is in order concerning the interpretation of the extended density regions between $(x, y) = (-3, 0)$ kpc and $(x, y) = (3, 0)$ kpc in both maps as evidence for a Galactic bar. We note in passing that observations of external barred spirals oftentimes show evidence for an alignment of the molecular bar (if present) with a stellar bar. Examples are M 83 (Lundgren et al. 2004; Hirota et al. 2018), M 100 (Sempere & Garcia-Burillo 1997), NGC 4569 (Boone et al. 2007), and even doubled-barred...
spirals such as NGC 3504 (Kuno et al. 2000; Wu et al. 2021). For the Milky Way, determinations of the bar inclination angle indicate values between 20° (Queiroz et al. 2021) and 45° (Anders et al. 2019), but neither of these angles agrees with the vanishing inclination angle of the feature described above. We therefore hold our verdict as to the physicality of this feature. Some N-body simulations do not show the formation of a molecular bar associated with the stellar bar (e.g. Athanassoula et al. 2013).

We note that the reconstructed gas density comprises a rather wide dynamical range, as expected for a log-normal density. Some of the fainter features are therefore difficult to identify on a linear scale. In order to highlight them, we show the mean of the posterior, again projected onto the Galactic plane on a logarithmic colour scale in Fig. 6. We also overlayed a longitude grid and the contours of the respective radial velocities. This should facilitate identifying a number of features with the corresponding features in the $\ell - v$ diagram of Fig. 2.

However, the mean of the posterior alone can be misleading because some of the localised features also have a rather large uncertainty. Unlike the previous deprojections, however, we now have a means of judging the validity of certain features by comparing the mean of the posterior with its uncertainty $\sigma$. To this end, we defined the signal-to-noise ratio (S/N) as $\mu/\sigma$. We show the S/N in Fig. 7, again for the BEG03 model in the top panel and in the bottom panel for the SBM15 model. Localised emission with an S/N of 3 or higher is clearly visible. In Fig. 7 we also overlaid the spiral arms, as determined from fits to a set of $\sim$200 masers (Reid et al. 2019; see their Table 2 for the fitted spiral parameters.). Many of the local emission features obtained for either gas-flow model can be easily associated with a spiral arm: for the BEG03 model for all spiral arms, but most impressively for the Norma, Sagittarius-Carina, Local, and Perseus arms. We comment on several noteworthy differences and similarities between the significant features obtained for the BEG03 and the SBM15 models below.

- The gas density in the SBM15 model is generally more scattered and does not cluster in regions as large as the emission in the BEG03 model. This is again due to the presence of local extrema in the radial velocity field in the BEG03 model, which boost the clustering. These local extrema are all but absent in the SBM15 model, and hence the gas density is less clustered.

- Some of the spiral arms are obvious also for the SBM15 model, however, for instance, the segments along the Perseus and Sagittarius-Carina arms for galactocentric azimuths $\varphi$ between $-200^\circ$ and $-280^\circ$. Other examples are the segments along the Norma arm ($90^\circ \leq \varphi \leq 150^\circ$) and the local arm ($330^\circ \leq \varphi \leq 0^\circ$).

- Some emission, in particular beyond the solar circle, is placed at different distances in the BEG03 and SBM15 models because of the different rotation curves adopted here. Given the rather small velocity gradient, this easily translates into differences of about a kiloparsec and thus affects the association with spiral arms. One example is emission around $\ell \sim 110^\circ$ and with $v_{lsr}$ between $-60$ and $-50 \, \text{km s}^{-1}$, see the bottom panel in Fig. 2. With the BEG03 model, this emission is located around $(x, y) = (10, -6) \, \text{kpc}$. With the SBM15 gas flow, this instead lies at $(x, y) = (9.5, -5) \, \text{kpc}$. In the former case, an association with the Norma arm suggests itself, in the latter case, the association with the Perseus arm is more likely.

In Fig. 8 we revisit the mass surface densities obtained with either gas-flow model (top panels) and compare them with the deprojections of Nakanishi & Sofue (2006) (bottom left) and Pohl et al. (2008) (bottom right). All gas densities are shown with the same dynamical range, but the mass surface density of Nakanishi & Sofue (2006) is significantly smoother. Some similarities are apparent between the map of Pohl et al. (2008) and our maps. This is even more so the case for our map that is based on the BEG03 model, the same gas-flow model as adopted by Pohl et al. (2008). However, there are also some differences:

- The spiral arms are more homogeneous in Pohl et al. (2008). While in our reconstruction, the width is oftentimes varying along the segments, in the gas density of Pohl et al. (2008), some ring segments appear to have constant width. This might be an artefact of the particular algorithmic reconstruction chosen there.

- An artefact that affects both the Pohl et al. (2008) map and our SBM15 map is the emission that is concentrated along an arc of the tangent point circle (see the discussion above).

- The peak in the gas density reconstructed for both our gas-flow models near $(x, y) = (-9, 2) \, \text{kpc}$ is absent in the Pohl et al. (2008) map, possibly because of the chosen suppression at large galactocentric radii and their enforcing of a vertical profile. We also note that our reconstruction places this feature at a rather large distance from the plane ($\sim 200 \, \text{pc}$). Furthermore, an inspection of the $v - \ell$ diagram shows that this feature has a very narrow line width, typical of local clouds with low masses. A similar conclusion might apply to the feature elongated between $(x, y) = (-6, 2) \, \text{kpc}$ and $(x, y) = (-6, 4) \, \text{kpc}$ in the BEG03 reconstruction.

- Another class of artefacts that is apparent in the Pohl et al. (2008) maps is the features that are elongated along the lines of sight, the fingers-of-god, especially in regions of low gas densities. These are almost completely absent in our reconstructions.

- Finally, Pohl et al. (2008) showed some gas density in regions where we found (almost) none, in particular beyond the solar circle. Unlike the structures that are solidly identified with spiral arms of the BEG03 model, some of these might be statistically not significant. This is to be compared with emission seen for our reconstruction in Fig. 6 near $(x, y) = (-14, -1) \, \text{kpc}$ and $(x, y) = (-14, -4) \, \text{kpc}$, which is not statistically significant, see Fig. 7.

We conclude our discussion considering some properties of the derived distributions in galactocentric radius. In Fig. 9 we plot the surface mass density averaged in galactocentric rings $R < R_i < R_{i+1}$, where $R_i = \Delta R$ with $\Delta R = 0.5 \, \text{kpc}$. Close to the Galactic centre, the average gas density is a few tens of $M_\odot \, \text{pc}^{-2}$ and it subsequently decreases towards $r = 3 \, \text{kpc}$. Beyond, there is a local maximum at $r = 5 \, \text{kpc}$, the well-established molecular ring (or possibly a convergence of spiral arms; Dobbs & Burkert 2012). Here, we find the average gas density to be $\sim 1 \, M_\odot \, \text{pc}^{-2}$ Beyond, the gas density decreases further. While it appears to reach a plateau of $2 \times 10^{-2} \, M_\odot \, \text{pc}^{-2}$ around 12 kpc, we recall that we did not detect much significant emission beyond $r \approx 10 \, \text{kpc}$.

We fitted Gaussian profiles $g(z) \propto \exp[-(z - z_\odot)^2/2\sigma_i^2]$ with means $z_\odot$ and standard deviations $\sigma_i$ to the gas density distribution in the same galactocentric rings $i$ as before by minimising the $\chi^2$ over all rings,

$$\chi^2 = \sum_i \sum_y \left[ \frac{\langle \Delta H_\odot, \odot y \rangle \odot \Delta r_i - g(z_y) \rangle}{\sigma_i^2} \right]$$

where $\langle \Delta H_\odot, \odot y \rangle \odot \Delta r_i$ denotes the averaging over all grid points $\alpha \beta$ in the $i$th ring. We fixed the uncertainty $\sigma_i$ to 0.15 of the maximum of $\langle \Delta H_\odot, \odot y \rangle \odot \Delta r_i$ for that particular ring $i$. For most rings, the Gaussian
Fig. 6. Projected mean gas density $\Sigma$ on a logarithmic colour scale, overlaying both contours in $v_{\text{LSR}}$ and a grid in longitude. Top: for the BEG03 gas flow model. Bottom: for the SBM15 model.
profile is a fair approximation, $\chi^2 < 1$. (Different $z$ are correlated, hence $\chi^2$ is usually $< 1$.) For other rings, however, the average profile in $z$ is only poorly described by a single Gaussian (see also Dame & Thaddeus 1985). We therefore removed these bins.

We show the full width at half maximum (FWHM), that is, $2 \sqrt{2 \log 2} \sigma$, a measure of the vertical extent of the gas distribution, as a function of galactocentric radius in the middle panel of Fig. 9 and compare it to a number of previous estimates. We note that our reconstructed gas densities appear to be more extended vertically, at least in the inner few kiloparsecs. Pohl et al. (2008) did not determine this parameter, but instead used it as an input for the analysis.

In the bottom panel of Fig. 9, we show the dependence of the midplane offset $z_0$ on the galactocentric radius and again...
compare it to a number of previous estimates. While the general trend with galactocentric radius is the same as seen in previous studies, our error bars are consistently smaller. We stress that the ensemble of samples from the posterior consists of full three-dimensional distributions, including uncertainties at individual positions and correlations between different positions. This enables statistical analysis beyond the axisymmetric quantities and simple error estimates.

4. Summary

We have presented a new deprojection of the CO line survey of Dame et al. (2001) with an unprecedented spatial resolution of 62.5 pc. This is based on a Gaussian variational Bayesian inference, which allows exploring the posterior distribution of this high-dimensional inference problem. While we assumed that correlations in configuration space exist, we did not assume any particular power spectrum, but determined the power spectrum during the reconstruction. We considered two gas-flow models that both take into account the presence of the Galactic bar, one based on a simulation of the gas flow in a predetermined potential (Bissantz et al. 2003, hereafter BEG03), the other based on a model for gas carrying orbits in the bar potential (Sormani et al. 2015, hereafter SBM15).

Our results are the three-dimensional distribution of molecular gas, assuming a fixed $X_{\text{CO}}$ factor of $2 \times 10^{20}$ molecules $\text{cm}^{-2} (\text{K} \text{km s}^{-1})^{-1}$. We have made our mean gas maps and their uncertainty available to the community. We showed the mean and standard deviation of the gas density projected onto the Galactic plane and compared with previous studies. Unlike earlier studies, we have the capacity of distinguishing between statistically significant structures and noise artefacts. We found that some of the most prominent structures are affected by the assumed spiral structure in the BEG03 gas-flow model, but that significant coherent structures, some of which align with spiral arms (e.g. defined by parallax measurements of masers) are present in the SBM15 model as well. We projected some radial profiles out of the three-dimensional gas distribution and largely confirm the results of previous studies.

In the future, several extensions of the current analysis are noteworthy. After successfully applying our method to molecular line surveys to determine the H$_2$ density, it would be interesting to apply it to surveys of atomic hydrogen, such as the recently completed HI4PI survey (HI4PI Collaboration 2016). While the HI distribution is known to show less clustering than molecular hydrogen, we are optimistic that the advantages of our approach will apply here as well. In addition, several shortcomings of the
Acknowledgements. We thank the anonymous referee for her/his constructive comments that helped improve the quality of this paper.

References

Ferrière, K. M. 2001, Rev. Mod. Phys., 73, 1031
Hastings, W. K. 1970, Biometrika, 57, 97
Kewley, L. J., Nicholls, D. C., & Sutherland, R. S. 2019, ARA&A, 57, 511
Kolmogorov, A. 1941, Akademija Nauk USSR Doklady, 30, 301

Fig. 9. Radial profiles of the surface mass density $\Sigma_{H_2}$ (top panel), FWHM of the $z$-profile (middle panel), and midplane offset (bottom panel). We compare the results for our two gas flow models (BEG03 and SBM15) with the results of earlier studies: Grabelsky et al. (1987), Bronfman et al. (1988), Nakanishi & Sofue (2006), Digel (1991), and Malhotra (1994).

present study are due to our ignorance of the gas flow. If the velocity field could be determined at the same time as the gas density, these deficiencies could be remedied. We can envisage two ways to help regularise this underdetermined problem: Data from the parallax measurements (Reid et al. 2019) could be included, thus providing at least some loci to which the velocity field could be anchored. In addition, the physical correlation between gas densities and flow velocities could link the reconstruction of the two fields, which would require additional model inputs, however.
Appendix A: Semi-analytical gas-flow model

Short of running our own hydrodynamical simulations of gas flow in a barred potential, we employed a semi-analytical approximation to the gas flow. It has been hypothesised by Binney et al. (1991) that gas in the potential of a rotating bar slowly drifts towards the centre, moving on orbits from two classes of closed orbits, so-called \( x_1 \) and \( x_2 \) orbits. The transition from the outer \( x_1 \) orbits to the inner \( x_2 \) orbits takes place through a shock structure that forms along a critical cusped orbit. Beyond this, the \( x_1 \) orbits self-intersect, but the gas instead drifts on \( x_2 \) orbits. This picture was observed in early simulations (Athanassoula 1992), but with limited resolution and a somewhat different potential than assumed by Binney et al. (1991), and more recently confirmed for the same potential and with a high-resolution simulation (SBM15). We adopted the potential of SBM15, but slightly adjusted their parameters.

Specifically, we adopted a combination of a triaxial potential

\[
\rho(r') = \frac{\rho_0}{4\pi a^3} \left( \frac{r'/a}{1 + r'/a} \right)^{\alpha - \beta},
\]

with \( r' = \sqrt{x^2 + (y/a)^2 + (z/a)^2} \) and a razor-thin disk potential generated by the surface mass density,

\[
\Sigma(R) = \Sigma_0 e^{-R/R_*},
\]

\( R \) denoting the cylindrical galactocentric radius. We adopted the following parameters: \( \rho_0 = 0.69 M_\odot \text{pc}^{-3} \), \( a = 1.8 \text{kpc} \), \( \alpha = 1.75 \), \( \beta = 3.5 \), \( \Sigma_0 = 1.3 \times 10^3 M_\odot \text{pc}^{-2} \), and \( R = 4.5 \text{kpc} \) (SBM15). We assumed a pattern speed of \( \Omega_p = 63 \text{ km s}^{-1} \text{kpc}^{-1} \) for the rigidly rotating bar potential.

We solved the equations of motion for test particles in the combined potential using the galpy package (Bovy 2015). We find closed \( x_1 \) and \( x_2 \) orbits in the range of galactocentric radius \( \leq 4 \text{kpc} \). Some examples are shown in Fig. A.1.

We interpolated the velocity field between the points on the \( x_1 \) and \( x_2 \) orbits for galactocentric radii \( R \leq 4 \text{kpc} \). While this works well for the regions inside the largest populated \( x_2 \) orbit and outside the smallest populated \( x_1 \) orbit, in between these orbits, the gas velocities are somewhat overestimated, see Fig. 5 of SBM15. We therefore added several nodes in the interpolation of the radial gas velocities to reach better agreement with Fig. 5 of SBM15.

Beyond \( 4 \text{kpc} \), we assumed gas to move on circular orbits and the velocity to follow the “universal rotation curve” of Persic et al. (1996) with the updated parameters of Reid et al. (2019). A small suppression of 5\% was necessary to cause the rotation curve to connect smoothly with the interpolation at \( R = 4 \text{kpc} \).

In the middle panel of Fig. 3 we show the radial velocity resulting from this semi-analytical gas-flow model. Even though we adopted slightly different parameters than SBM15, we refer to this model as the SBM15 model in the main text.

\[\text{http://github.com/jobovy/galpy}\]