



Universiteit
Leiden
The Netherlands

exoALMA. VIII. Probabilistic moment maps and data products using nonparametric linear models

Hilder, T.; Casey, A.R.; Price, D.J.; Pinte, C.; Izquierdo, A.F.; Hardiman, C.; ... ; Zawadzki, B.

Citation

Hilder, T., Casey, A. R., Price, D. J., Pinte, C., Izquierdo, A. F., Hardiman, C., ... Zawadzki, B. (2025). exoALMA. VIII. Probabilistic moment maps and data products using nonparametric linear models. *Astrophysical Journal Letters*, 984(1). doi:10.3847/2041-8213/adc435

Version: Publisher's Version

License: [Creative Commons CC BY 4.0 license](#)

Downloaded from: <https://hdl.handle.net/1887/4290588>

Note: To cite this publication please use the final published version (if applicable).



exoALMA. VIII. Probabilistic Moment Maps and Data Products Using Nonparametric Linear Models

Thomas Hilder¹ , Andrew R. Casey^{1,2,3} , Daniel J. Price¹ , Christophe Pinte⁴ , Andrés F. Izquierdo^{5,6,7,27} , Caitlyn Hardiman¹ , Jaehan Bae⁷ , Marcelo Barraza-Alfaro⁸ , Myriam Benisty^{9,10} , Gianni Cataldi¹¹ , Pietro Curone^{12,13} , Ian Czekala¹⁴ , Stefano Facchini¹² , Daniele Fasano⁹ , Mario Flock¹⁰ , Misato Fukagawa¹¹ , Maria Galloway-Sprietsma⁷ , Himanshi Garg¹ , Cassandra Hall^{15,16,17} , Iain Hammond¹ , Jane Huang¹⁸ , John D. Ilee¹⁹ , Kazuhiro Kanagawa²⁰ , Geoffroy Lesur⁴ , Cristiano Longarini^{12,21} , Ryan Loomis²² , Ryuta Orihara²⁰ , Giovanni Rosotti¹² , Jochen Stadler^{4,9} , Richard Teague⁸ , Hsi-Wei Yen²³ , Gaylor Wafflard⁴ , Andrew J. Winter^{9,10} , Lisa Wölfer⁸ , Tomohiro C. Yoshida^{11,24} , and Brianna Zawadzki^{25,26}

¹ School of Physics and Astronomy, Monash University, VIC 3800, Australia; thomas.hilder@monash.edu

² Center for Computational Astrophysics, Flatiron Institute, 162 5th Avenue, NY, USA

³ Centre of Excellence for Astrophysics in Three Dimensions (ASTRO-3D), Melbourne, VIC, Australia

⁴ University Grenoble Alpes, CNRS, IPAG, F-38000 Grenoble, France

⁵ Leiden Observatory, Leiden University, P.O. Box 9513, NL-2300 RA Leiden, The Netherlands

⁶ European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching bei München, Germany

⁷ Department of Astronomy, University of Florida, Gainesville, FL 32611, USA

⁸ Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁹ Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, France

¹⁰ Max-Planck Institute for Astronomy (MPIA), Königstuhl 17, 69117 Heidelberg, Germany

¹¹ National Astronomical Observatory of Japan, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

¹² Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy

¹³ Departamento de Astronomía, Universidad de Chile, Camino El Observatorio 1515, Las Condes, Santiago, Chile

¹⁴ School of Physics & Astronomy, University of St. Andrews, North Haugh, St. Andrews KY16 9SS, UK

¹⁵ Department of Physics and Astronomy, The University of Georgia, Athens, GA 30602, USA

¹⁶ Center for Simulational Physics, The University of Georgia, Athens, GA 30602, USA

¹⁷ Institute for Artificial Intelligence, The University of Georgia, Athens, GA 30602, USA

¹⁸ Department of Astronomy, Columbia University, 538 W. 120th Street, Pupin Hall, New York, NY 10027, USA

¹⁹ School of Physics and Astronomy, University of Leeds, Leeds, LS2 9JT, UK

²⁰ College of Science, Ibaraki University, 2-1-1 Bunkyo, Mito, Ibaraki 310-8512, Japan

²¹ Institute of Astronomy, University of Cambridge, Madingley Rd, CB30HA, Cambridge, UK

²² National Radio Astronomy Observatory, Charlottesville, VA 22903, USA

²³ Academia Sinica Institute of Astronomy & Astrophysics, 11F of Astronomy-Mathematics Building, AS/NTU, No. 1, Sec. 4, Roosevelt Rd., Taipei 10617, Taiwan

²⁴ Department of Astronomical Science, The Graduate University for Advanced Studies, SOKENDAI, 2-21-1 Osawa, Mitaka, Tokyo 181-8588, Japan

²⁵ Department of Astronomy, Van Vleck Observatory, Wesleyan University, 96 Foss Hill Drive, Middletown, CT 06459, USA

²⁶ Department of Astronomy & Astrophysics, 525 Davey Laboratory, The Pennsylvania State University, University Park, PA 16802, USA

Received 2024 November 21; revised 2025 March 5; accepted 2025 March 8; published 2025 April 28

Abstract

Extracting robust inferences on physical quantities from disk kinematics measured from Doppler-shifted molecular line emission is challenging due to the data's size and complexity. In this paper, we develop a flexible linear model of the intensity distribution in each frequency channel, accounting for spatial correlations from the point-spread function. The analytic form of the model's posterior enables probabilistic data products through sampling. Our method debiases peak intensity, peak velocity, and line width maps, particularly in disk substructures that are only partially resolved. These are needed in order to measure disk mass, turbulence, and pressure gradients and detect embedded planets. We analyze HD 135344B, MWC 758, and CQ Tau, finding velocity substructures 50–200 m s⁻¹ greater than with conventional methods. Additionally, we combine our approach with DISCMINER in a case study of J1842. We find that uncertainties in stellar mass and inclination increase by an order of magnitude due to the more realistic noise model. More broadly, our method can be applied to any problem requiring a probabilistic model of an intensity distribution conditioned on a point-spread function.

Unified Astronomy Thesaurus concepts: [Astronomy data modeling \(1859\)](#); [Nonparametric inference \(1903\)](#); [Linear regression \(1945\)](#); [Radio interferometry \(1346\)](#); [Protoplanetary disks \(1300\)](#); [Planet formation \(1241\)](#); [Planetary-disk interactions \(2204\)](#)

1. Introduction

Observations of circumstellar disks in Doppler-shifted molecular line emission at high spatial and spectral resolution, as obtained in the exoALMA project (R. Teague et al. 2025), offer a wealth of information about the 3D velocity field (A. F. Izquierdo et al. 2025), emitting-layer height, temperature structure (M. Galloway-Sprietsma et al. 2025), rotation curve, star and disk masses (C. Longarini et al. 2025), radial

²⁷ NASA Hubble Fellowship Program Sagan Fellow.



kinematic substructures (J. Stadler et al. 2025), turbulence (M. Barraza-Alfaro et al. 2025), and spiral arms due to planets or stellar-mass companions (C. Pinte et al. 2025).

Performing these measurements in a statistically robust way is difficult. First, the data are complicated. At the high spatial resolutions achieved in exoALMA, we resolve many local nonaxisymmetric features in both intensity and velocity (A. F. Izquierdo et al. 2025), which cannot be explained with global axisymmetric Keplerian disk models. While Bayesian approaches modeling either channel maps (A. F. Izquierdo et al. 2021, 2022, 2023) or visibilities (I. Czekala et al. 2015; K. M. Flaherty et al. 2015; K. Flaherty et al. 2020; J. Pegues et al. 2021; N. T. Kurtovic & P. Pinilla 2024) with parametric disk models are able to infer all parameters simultaneously, their lack of flexibility limits the interpretability of the posterior (e.g., J. Bernardo & A. Smith 2009) when applied to disks containing substructures as seen in all of the exoALMA targets (Disk Dynamics Collaboration et al. 2020; C. Pinte et al. 2025; R. Teague et al. 2025).

Second, the imaging process creates correlations between image pixels, and the finite sampling of the uv -plane (e.g., S. M. Andrews et al. 2018) means that the data contain no information about particular frequencies. The resultant image point-spread function results in the underestimation of line intensity gradients known as beam smearing (W. D. Cotton 1989). Beam smearing is known to bias line widths in the inner disk (R. Teague et al. 2016) and velocities in the presence of intensity gradients, including the local gradients caused by substructures (M. Keppler et al. 2019; Y. Boehler et al. 2021). It also causes bias in rotation curves even at radii greater than 4 times the full width at half-maximum (FWHM) of the beam (S. M. Andrews et al. 2024; J. Stadler et al. 2025). Recently, S. M. Andrews et al. (2024) found empirical corrections for this effect for rotation curve analysis, but such corrections are limited to axisymmetric Keplerian disks, and local intensity gradients are not accounted for.

Third, analyses looking to measure deviations from Keplerian rotation, such as from pressure gradients (R. Teague et al. 2018; G. P. Rosotti et al. 2020; H. Garg et al. 2022; A. F. Izquierdo et al. 2022, 2023; J. Stadler et al. 2025), self-gravity of the gas (B. Veronesi et al. 2021; G. Lodato et al. 2023; C. Longarini et al. 2025; P. Martire et al. 2024; J. Speedie et al. 2024), winds (M. Galloway-Sprietsma et al. 2023), and meridional flows (R. Teague et al. 2019), often fix geometric parameters such as inclination and position angle, as well as the emitting-layer height and temperature distributions, even though these quantities themselves are uncertain. This issue is addressed simply either with Monte Carlo error propagation (S. M. Andrews et al. 2024) or with bootstrapping (M. Galloway-Sprietsma et al. 2025), provided there is some reasonable quantification of the uncertainty on those parameters.

In this paper, we present a first step toward building a method that can in principle address *all* of the above concerns. We provide a flexible, nonparametric model for intensity distribution in each channel of a spectral cube, including the convolution with the beam in the data generation process, as well as the correlations created between neighboring pixels. Samples from the posterior may be generated cheaply due to the analytic posterior and fed through any cube analysis pipeline to calculate uncertainties and correlations in measurements. We demonstrate a simple example use case of the model in the generation of probabilistic peak intensity, peak velocity,

and line width maps that are unbiased by beam smearing. We also combine our approach with DISCMINER (A. F. Izquierdo et al. 2021, 2022, 2023) to investigate the effect of the more realistic noise model, including spatial correlations, on the precision of the inferred disk model parameters.

The paper is organized as follows. We present the model and analysis methods in Section 2, and we benchmark the method in Section 3. Section 4 presents our application of the model to a subset of the exoALMA sample, while Section 5 discusses our main findings. We conclude in Section 6.

2. Methods

2.1. Data

We perform all analysis using the fiducial $^{12}\text{CO } J=3-2$ cubes (R. Teague et al. 2025) for HD 135344B, MWC 758, CQ Tau, and J1842. These have a circular synthesized beam with an FWHM of $0''.15$ and a channel spacing of 100 m s^{-1} .

HD 135344B, MWC 758, and CQ Tau are used to demonstrate the generation of probabilistic peak intensity, peak velocity, and line width maps using our method. These sources were chosen due to their low inclination, allowing us to treat the line profile as single-peaked due to their lack of backside contamination and because they all feature large spiral features in CO (Y. Boehler et al. 2018; S. Casassus et al. 2021; L. Wölfer et al. 2021; A. F. Izquierdo et al. 2025). We emphasize that the method does not ultimately require any particular line profile shape but that this choice allows us to focus on testing downstream data products in a relatively simple setting.

J1842 is used for the case study combining our method with DISCMINER and was chosen as a representative typical source from the sample. By this, we mean that it is mid-inclination (39°) and orbits an approximately $1 M_\odot$ star (R. Teague et al. 2025) and that the channels feature only modest deviations from smooth rotation (C. Pinte et al. 2025). The ^{12}CO is also only $5''$ across, which aids in computational cost.

We downsample each cube by a factor of 2 in each spatial axis, replacing 2×2 blocks of pixels with their mean value. The cubes are also cropped spatially to include only the region where the signal is contained. Both these steps are done to reduce computational expense, which is discussed further in Section 5.

2.2. Nonparametric Model

We elect to directly model the intensity distribution in each channel. While the model for each channel is independent of the others, for the sake of simplicity, we will refer to the collection of model channels comprising the whole cube as a “model cube.” The model is nonparametric and consists of a linear summation of orthogonal basis functions, where the coefficients or *weights* of these functions are determined by fitting the model to the data. This *linear* approach is crucial for allowing for tractable sampling from the model posterior distribution, even with of order 10^4 model parameters, as the posterior takes the form of a multivariate normal distribution from which we can sample in a few minutes.

Let \mathbf{I} be a vector containing the model intensities I_i from a particular channel, where i is an index over the $n = n_x n_y$ pixels in the channel. We model I_i as a weighted sum of p 2D basis functions,

$$I(x_i, y_i) = \sum_{j=1}^p X_j f_j(x_i, y_i), \quad (1)$$

where x_i and y_i are spatial coordinates for pixel i and X_j is the weight associated with basis function f_j . We choose to use a Fourier series for the basis, which consists of orthogonal trigonometric functions in both the x - and y -directions. We expand on this choice later in this section. This allows us to decompose f_j into two identical sets of basis functions, one for each spatial direction, giving

$$I(x_i, y_i) = \sum_{l=1}^{p_x} \sum_{m=1}^{p_y} X_{lm} [g_l(x_i) \cdot g_m(y_i)], \quad (2)$$

where $p = p_x p_y$, $l = j \bmod p_x$, $m = \lfloor j/p_x \rfloor$, $f_j(x, y) = g_l(x) \cdot g_m(y)$, and $\lfloor \cdot \rfloor$ denotes the floor function. The basis functions g_l are defined as (for both l and m)

$$g_l(x) = \begin{cases} \cos(\omega_l x) & \text{for } l \text{ odd} \\ \sin(\omega_l x) & \text{for } l \text{ even} \end{cases}, \quad \omega_l = \frac{\pi}{L} \left\lfloor \frac{l}{2} \right\rfloor, \quad (3)$$

where L is a length scale in the x -space. We choose spatial coordinates x_i, y_i for the pixels uniformly spaced between 0 and 1 and set $L = 3$ following D. W. Hogg & S. Villar (2021). By defining the entries of the $n_x \times p_x$ matrix A_x as

$$[A_x]_{ij} = g_j(x_i), \quad (4)$$

and similarly for A_y , we can rewrite Equation (2) simply as

$$I = (A_x \otimes A_y) X = A_{xy} X, \quad (5)$$

where X is a vector of length p containing the Fourier weights X_j , A_{xy} is an $n \times p$ matrix, and \otimes is the Kronecker product. We note that our choice of basis functions does not constrain the model intensities to be positive. While the physical intensity is strictly positive, the intensity in the synthesized images is not. Since we only have access to noisy measurements rather than the true underlying intensities, allowing the model to take nonpositive values is a deliberate and justified choice.

Let $h(x, y)$ be a function representing the kernel of the beam, a circular 2D Gaussian with FWHM = 0.15 for our purposes, but it may be of any form in general. Defining the entries of the $n \times n$ matrix H as

$$[H]_{ij} = h(x_i - x_j, y_i - y_j), \quad (6)$$

we can find the vector of the beam-convolved intensities I_{conv} from

$$I_{\text{conv}} = H I = H A_{xy} X = A X, \quad (7)$$

where the last equality is from defining the *design matrix* $A = H A_{xy}$.

Defining I_{data} as a vector containing the actual data observed from a particular channel, we can write our overall model as

$$I_{\text{data}} = A X + \epsilon, \quad (8)$$

where ϵ is a random vector of length n representing the noise in the image. We will assume that the noise follows a multivariate normal distribution (hereafter just normal distribution)

$$\epsilon \sim \mathcal{N}(\mathbf{0}, C), \quad (9)$$

where the \sim notation denotes that ϵ is drawn from the normal distribution \mathcal{N} with zero mean vector and $n \times n$ covariance matrix C . We assume that the noise is spatially correlated on

the scale of the beam, which gives

$$[C]_{ij} = \sigma_i \sigma_j [H]_{ij} / \max(H), \quad (10)$$

where σ_i is the marginal standard deviation of the noise in pixel i . In practice, we assume the same σ for all pixels in the cube, which we estimate from the standard deviation of the data in channels with no significant signal.

We emphasize that this approach of modeling the image intensities only, and *not* treating the finite sampling of the uv -plane, is a nontrivial simplification. Ideally, we would forward model the visibilities and include a treatment of the uv -sampling function; however, we leave this for future work.

The model as described so far constitutes a *linear regression model*, even though the basis functions themselves are nonlinear. The typical approach to fitting the above model would be to perform ordinary least squares (OLS), where one finds the “best” parameter vector X , which we will label \hat{X} , by minimizing the residuals squared,

$$\hat{X} = \operatorname{argmin}_X \|I_{\text{data}} - A X\|^2, \quad (11)$$

which may be solved via the matrix pseudoinverse²⁸ (E. H. Moore 1920; R. Penrose 1956),

$$\hat{X} = (A^T A)^{-1} A^T I_{\text{data}}. \quad (12)$$

This approach may be extended to include the covariance matrix and can be understood as a simple reweighting of the data by the inverse of their covariances,

$$\hat{X} = (A^T C^{-1} A)^{-1} A^T C^{-1} I_{\text{data}}, \quad (13)$$

which is called the generalized least squares (GLS) estimate. The GLS estimate has some desirable properties, namely, that it is the best linear unbiased estimator²⁹ and that it maximizes the model likelihood.

This approach will always result in a unique solution for \hat{X} and the reconstructed I_{conv} . However, we are interested in the model intensities prior to beam convolution I given by Equation (5), which becomes increasingly poorly behaved as the number of basis functions p is increased. Because H acts in the design matrix to dampen signal contributed from high-frequency Fourier modes, there is little information available to constrain the weights of those modes. I therefore has many possible values that result in similar I_{conv} . This problem is identical to the linear deconvolution problem in the presence of noise, which has been studied extensively for at least 80 yr (e.g., N. Wiener 1949; D. W. Oldenburg 1981).

A natural solution is to provide additional constraints on \hat{X} that will result in an I that we believe to be more sensible (e.g., A. Gelman et al. 2014). From a frequentist perspective, this additional constraint is called a regularizer, while the (equivalent) Bayesian view is that of a prior probability distribution over possible X . Here we will mostly adopt the language of the latter as we wish to use the posterior probability distribution (the resultant distribution from updating the prior with the data via Bayes’s theorem) in our analysis. We

²⁸ The matrix $A^T A$ is only invertible for $p < n$. We use $p > n$ in this paper, but this is not an issue as long as the regularization term $[\Lambda]_{jj}$ we introduce is strictly positive for all j . This is the case; see Equations (18) and (19).

²⁹ This is because GLS is equivalent to OLS under an affine transformation, and OLS is the best linear unbiased estimator via the Gauss–Markov theorem (e.g., M. Kutner et al. 2004).

first assume some generic normal prior on \mathbf{X} ,

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda}), \quad (14)$$

where $\boldsymbol{\mu}$ is a mean vector of length p and $\boldsymbol{\Lambda}$ is a $p \times p$ covariance matrix. The posterior distribution over \mathbf{X} has an analytic form thanks to the linearity of the model and because both the noise and prior are normal distributions,

$$p(\mathbf{X} | \mathbf{I}_{\text{data}}) = \mathcal{N}(\mathbf{X} | \hat{\mathbf{X}}, \boldsymbol{\Sigma}), \quad (15)$$

where the right-hand side is shorthand notation for the probability density of the normal distribution with mean vector $\hat{\mathbf{X}}$ and covariance matrix $\boldsymbol{\Sigma}$ evaluated at some \mathbf{X} . Additionally (e.g., D. W. Hogg et al. 2020),

$$\hat{\mathbf{X}} = \boldsymbol{\Sigma}(\mathbf{A}^\top \mathbf{C}^{-1} \mathbf{I}_{\text{data}} + \boldsymbol{\Lambda}^{-1} \boldsymbol{\mu}), \quad (16)$$

$$\boldsymbol{\Sigma} = (\mathbf{A}^\top \mathbf{C}^{-1} \mathbf{A} + \boldsymbol{\Lambda}^{-1})^{-1}, \quad (17)$$

where the differences between the above and Equation (13) can be understood as reweighting the predictions using the prior.

We choose a prior that penalizes high-frequency basis functions more than low-frequency ones in order to encourage smoothness. This can be achieved with feature weighting (B. Bah & R. Ward 2016; H. Rauhut & R. Ward 2016; Y. Xie et al. 2022), where we set $\boldsymbol{\mu} = 0$ and the prior variance for each basis function using some weighting function w that decreases with frequency,

$$[\boldsymbol{\Lambda}]_{jj} = \frac{1}{\lambda} [w(\omega_l, \omega_m)]^2, \quad (18)$$

recalling that ω_l and ω_m are the x - and y -direction frequencies for basis function f_j , and we choose (D. W. Hogg & S. Villar 2021)

$$w(\omega_l, \omega_m) = \left(\frac{\pi}{2} \right)^{1/4} \left(\frac{1}{s} + s[\omega_l^2 + \omega_m^2] \right)^{-1/2}, \quad (19)$$

where both λ and s are *hyperparameters* for our model that encode how strong our prior is. λ is an overall scale that uniformly scales the variance across all X_j , while s controls how quickly the variance decreases with the frequency of the basis. We somewhat arbitrarily choose values for the hyperparameters of $\lambda = s = 0.1$, since we found them to be performant in terms of predicting the data (see Section 3) without fine-tuning. We discuss the caveats associated with fixing the hyperparameters to arbitrary values, in particular the impacts on the resultant model uncertainties, in Appendix A.

We also choose the *overparameterized* case, where we have more model parameters than data points $p > n$. We set the number of basis functions in each direction p_x, p_y to be 1.25 times the number of pixels in each direction n_x, n_y , rounded to the nearest integer. We justify this choice by returning to our choice of basis. It was realized in the last few years that a Fourier basis is an effective way to represent stationary kernels for Gaussian process (GP) regression (J. Hensman et al. 2018; D. W. Hogg & S. Villar 2021; P. Greengard et al. 2022). GPs are distributions over functions (e.g., C. E. Rasmussen & C. K. I. Williams 2006). In the limit $p \rightarrow \infty$, our basis and prior choice is equivalent to a GP prior on the model intensities,

$$I(\mathbf{x}) \sim \mathcal{GP}(0, k(\mathbf{x}, \mathbf{x}')), \quad (20)$$

where $\mathbf{x} = [x, y]^\top$ and k is a function encoding the model covariance between two points, usually called the kernel.

Specifically, the equivalence is given by

$$\lim_{p \rightarrow \infty} \sum_{j=1}^p [\boldsymbol{\Lambda}]_{jj} f_j(\mathbf{x}) f_j(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}'), \quad (21)$$

and we refer the reader to J. Hensman et al. (2018) and D. W. Hogg & S. Villar (2021) for further details. F. Tobar et al. (2023) recently showed that GP priors can be used in this manner to solve the noisy deconvolution problem in a Bayesian setting, outperforming conventional approaches such as the Wiener filter while simultaneously providing uncertainties on the recovered signal. While our approach is only an approximation of theirs, kernels with rapidly decaying Fourier transforms are well approximated by an equispaced Fourier series for even relatively small p (P. Greengard et al. 2022). Our choice of weighting function given in Equation (19) is equivalent to the exponential kernel (C. E. Rasmussen & C. K. I. Williams 2006; D. W. Hogg & S. Villar 2021), and its Fourier transform decays quadratically.

2.3. Fitting and Sampling

We perform the fits by calculating the posterior mean $\hat{\mathbf{X}}$ and covariances $\boldsymbol{\Sigma}$ using Equations (16) and (17), respectively. Note that while $\hat{\mathbf{X}}$ must be calculated for each channel separately, $\boldsymbol{\Sigma}$ depends only on quantities that are identical for all channels and only needs to be calculated once for each cube. The best-fitting model channels are then generated as $\mathbf{I} = \mathbf{A}_{xy} \hat{\mathbf{X}}$, while the best-fitting convolved model is $\mathbf{I}_{\text{conv}} = \mathbf{A} \hat{\mathbf{X}}$. We additionally draw 100 posterior samples of \mathbf{X} for each channel from the posterior $\mathcal{N}(\hat{\mathbf{X}}, \boldsymbol{\Sigma})$ and convert them to posterior samples of \mathbf{I} and \mathbf{I}_{conv} in the same way as for the best fit. “Sample cubes” are assembled by grouping a unique sample of each channel together, yielding 100 model cube samples, both before and after beam convolution. Details about the numerical methods and accuracy considerations for both the fitting and sampling are given in Appendix B.

2.4. General Probabilistic Data Products

Figure 1 illustrates the three steps involved in applying our method generally to generate probabilistic data products from spectral cube data. They are as follows.

1. The model is fit to the data cube to determine the posterior distribution for each channel.
2. Posterior model cube samples are generated.
3. The data products of interest are calculated independently from each cube sample, yielding posterior samples of each product.

As shown in Figure 1 and indicated in Section 2.3, the posterior model cube samples can be generated as either beam-convolved (blue in Figure 1) or not (orange). This choice would generally depend on the specific interests of the investigator. In our analysis, we will use the unconvolved cube samples for the peak intensity, peak velocity, and line width maps. For the case study using DISCMINER, we use the convolved cube samples since DISCMINER forward models the effect of the beam.

The posterior data product samples can be used either to reconstruct the marginal posterior distribution over that product or just to find a most likely value with associated uncertainties.

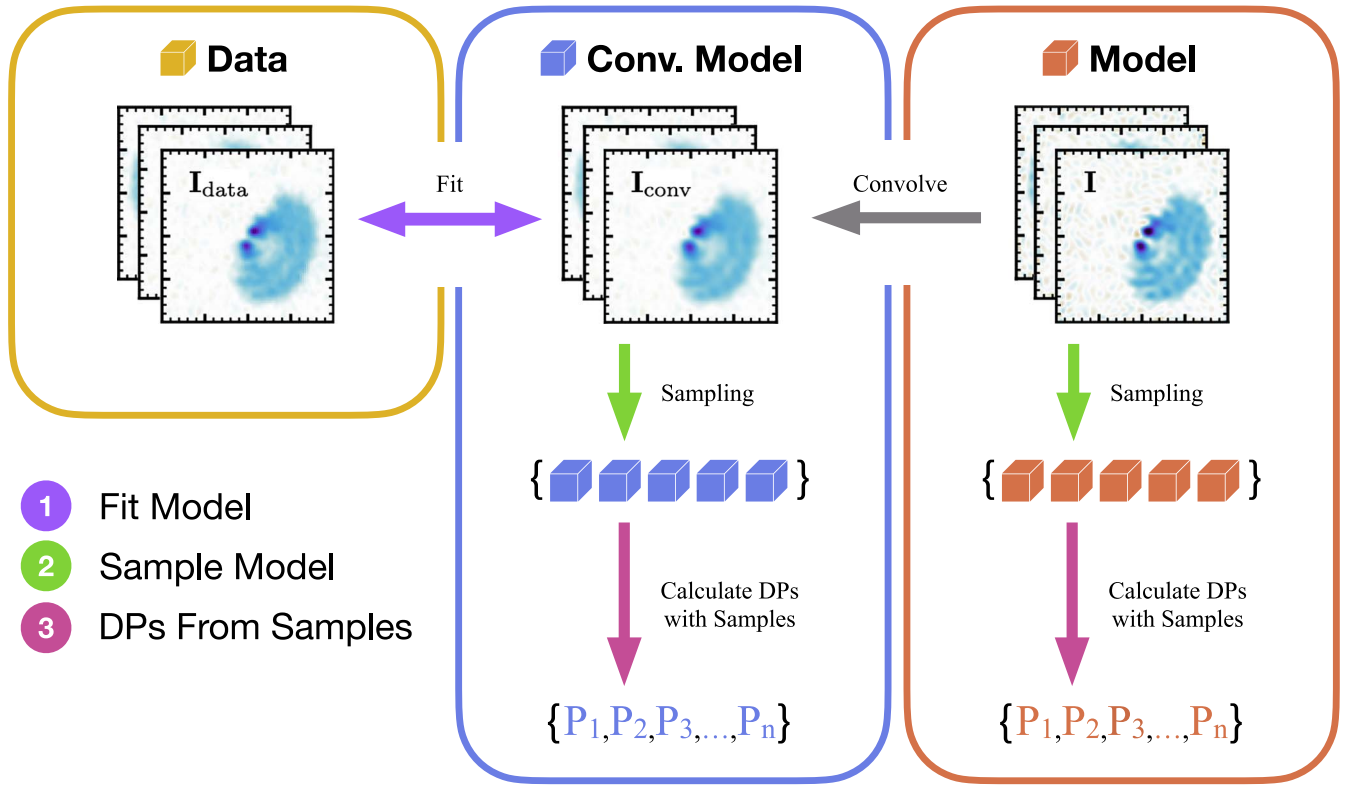


Figure 1. Schematic overview of the methodology for generating probabilistic data products using the nonparametric cube model. The cube graphic represents data cubes, with the orange and blue cubes illustrating posterior model cube samples before and after beam convolution, respectively. The P_i values indicate sample data products (DPs) calculated from the corresponding cube samples, color-coded accordingly. The numbered steps in the bottom left correspond directly to those in Section 2.4.

The distribution is estimated simply from a histogram or kernel density estimate, while the best value and uncertainties can be taken from the median or mean and chosen quantiles, respectively (e.g., D. W. Hogg & D. Foreman-Mackey 2018).

2.5. Probabilistic Moment Maps

As a simple example for a probabilistic data product, we generate peak intensity, peak velocity, and line width maps, commonly called moment maps, following our method. We do this for HD 135344B, CQ Tau, and MWC 758 as outlined in Section 2.1.

We fit Gaussian line profiles to each pixel of each sample cube

$$f(v_{\text{ch}}, \theta_0) = I_0 \exp\left(-\frac{1}{2} \left[\frac{v_{\text{ch}} - v_0}{w_0} \right]^2\right), \quad (22)$$

where $\theta_0 = [I_0, v_0, w_0]$, I_0 is the peak intensity, v_0 is the peak or centroid velocity, and w_0 is the line width. Each line profile in each cube sample will have its own best-fitting θ_0 , but we neglect indices over these in the notation for simplicity. We use maximal likelihood estimation to perform the fit, in practice minimizing the negative log likelihood given by

$$\log \mathcal{L}_m(\mathbf{I} | \theta_0) = -\frac{1}{2\sigma^2} \sum_i^{n_{\text{ch}}} [I_i - f(v_{\text{ch},i}, \theta_0)]^2, \quad (23)$$

where i is an index over channels, and we assume the noise σ is normally distributed. We use the L-BFGS-B optimization algorithm (C. G. Broyden 1970; R. Fletcher 1970; D. Goldfarb

1970; D. F. Shanno 1970; D. C. Liu & J. Nocedal 1989) implemented in SciPy’s `optimize.minimize` function (P. Virtanen et al. 2020) to fit. A noise cutoff of 3 times the cube rms was applied to avoid fitting lines with low signal-to-noise.

After applying the above procedure to each sample cube (preconvolution), we can estimate best fits and associated uncertainties over I_0 , v_0 , w_0 for each data cube, as described in Section 2.4 above. The best-fitting θ_0 for each line is taken as the median across the 100 cube samples, while the 1σ uncertainty is estimated from the standard deviation. We present these results in Section 4.1.

We also fit Gaussian line profiles to the data directly, as is typically done in studies of kinematics or other disk properties, for comparison. In our results, we refer to the probabilistic maps described above as the “new method” and the fit directly to the data as the “conventional method.” We emphasize that the important distinction is the use of the nonparametric model and sampling in our method over the direct fit to the data in the conventional method.

2.6. DISCMINER Case Study

DISCMINER is a statistical inference tool for protoplanetary disk properties that uses a forward model of line emission to constrain global properties of the disk including the mass of the central star, the inclination and position angle, and the radial height and intensity profiles (A. F. Izquierdo et al. 2021). DISCMINER is used extensively in the analysis performed by exoALMA (A. F. Izquierdo et al. 2025), as the parametric disk model it provides is ideal for quantifying localized deviations

in, for example, the rotation curve (C. Longarini et al. 2025; J. Stadler et al. 2025) or pressure scale height (J. Stadler et al. 2025).

DISCMINER uses Bayesian inference to estimate the posterior distribution of the disk model, allowing for quantification of the model uncertainties. However, it uses a simplified noise model where the correlations between nearby pixels in the cube are neglected. This simplification is to avoid the increased computational expense incurred when evaluating a multivariate normal likelihood during the Markov Chain Monte Carlo sampling. Such correlations are known to widen the posterior distribution, so their exclusion will in general result in underestimated uncertainties (e.g., A. Gelman et al. 2014), and predicting the factor by which they are underestimated is not straightforward.

Our nonparametric model of the channels can cheaply incorporate spatial correlations in the noise because it is linear (see Section 2.2). This allows us to combine our approach with DISCMINER to assess the impact of these correlations on inference precision. We use J1842 as a test case for this combined approach.

To prevent ambiguity, we will refer to posterior samples from our nonparametric model as “cube samples” and posterior samples from the DISCMINER model as “DISCMINER samples.” We generated 30 cube samples according to the procedure outlined in Section 2.4, in this case using the *convolved* samples as DISCMINER forward models the effect of the beam. We fit each cube sample with DISCMINER, obtaining many posterior distributions, where each is conditioned on a single cube sample. Finally, we estimate the combined posterior by aggregating the posterior DISCMINER samples from all fits.

Additionally, we also refit DISCMINER to the fiducial J1842 data cube downsampled by the same factor as used for the nonparametric model fit (see Section 2.1). The production DISCMINER fit for J1842 used across the other exoALMA papers (A. F. Izquierdo et al. 2025; R. Teague et al. 2025) used downsampling of 6×6 pixel blocks, whereas we used 2×2 . We compare these results with the combined approach in Section 4.2.

3. Benchmark

3.1. Synthetic Data and Fit

For the benchmark model, we assume that the line in each pixel is Gaussian and choose radial intensity, velocity, and line width profiles, as well as other model parameters, to create similar-looking synthetic data to HD 135344B. The specific radial profiles we choose are given in Appendix C. For the model, we use central mass $M_* = 1.61 M_\odot$, inclination $i = 16.11^\circ$, position angle $PA = 242.94^\circ$ (A. F. Izquierdo et al. 2025), and a distance of 135 pc (Gaia Collaboration et al. 2023). The true velocity, intensity, and line width maps $v_{0,\text{truth}}$, $I_{0,\text{truth}}$, and $w_{0,\text{truth}}$ are calculated from the radial profiles by rotating sky coordinates to match the inclination and position angle of the model and for the velocities by taking only the line-of-sight component.

True channels are created using the true maps using the Gaussian line profile in Equation (22), with v_{ch} in increments of 100 m s^{-1} . These channels are shown in the fourth row of Figure 2. We then convolved the channels spatially with a circular Gaussian beam with $\text{FWHM} = 0''.15$ to match the data

and added beam-correlated noise with scale $\sigma = 20 \text{ mJy}$. These synthetic data are shown in the top row of Figure 2.

We fit the synthetic data using the nonparametric model. The best-fitting model and convolved model are shown in the third and fourth rows of Figure 2. The residuals between the synthetic data and the fit (fifth row) are all on the scale of the noise or less, with the most significant difference found near the center of the disk. By comparing the unconvolved model fit with the true channels and with the data, we see that the model is capable of recovering the steep intensity gradients colocated with the substructures that are partially smeared out by the beam. Similarly, the intensity of the channels near the center is better resolved in the model compared to the data, although the model is not able to reconstruct the very sharp edges present in the true channels in this region. The latter effect is likely explained by the model prior, which encourages smoothness. The last row of the figure shows the residuals between the model and true channels, which are mostly just noise except in the inner regions. This indicates that the model effectively provides an unbiased estimate of the intensity, except within the central few beams.

Note also that the model *contains noise*. Due to its nonparametric nature, the model effectively does not distinguish between signal and noise in the data. Indeed, the model before beam convolution is noisier than the data itself due to the uncertainty associated with the information lost through beam convolution.

3.2. Moment Maps and Substructures

The middle columns of Figures 3 and 4 present the probabilistic v_0 , I_0 , and w_0 maps calculated following our new method. These may be compared to maps calculated with the conventional method (see Section 2.5), shown in the left columns. The differences between each method and the true map used for the synthetic data are shown in the bottom rows. The top-right panels show $v_{0,\text{truth}}$, $I_{0,\text{truth}}$, and $w_{0,\text{truth}}$, respectively.

Comparing the peak velocity maps (Figure 3, top plot), we can see from the bottom panels that the new method provides a less biased estimate of the velocity along the ring substructures. As expected, the conventional method is biased by beam smearing, resulting in the underestimation of the velocity perturbations of $50\text{--}200 \text{ m s}^{-1}$. This bias is not present in our new method. This effect can also be seen in the contours drawn in the top panels, which are commonly used to trace disk substructures (e.g., J. Calcino et al. 2022). The “wiggles” in the contours of the conventional method are smaller than in the true map, while those in the new method map more accurately match the truth. Both methods underestimate the velocities in the central beam or so, although the new method performs better as the biased region is smaller. The new method contains some artifacts aligned with the vertical and horizontal directions, which could be due the 2D Fourier basis causing a preferred direction in the model.

The peak intensity maps (Figure 3, bottom plot) show that the conventional map systematically underestimates the peak intensity in most of the disk, although only by around 10%. The new method does not contain this bias throughout most of the disk. The central few beams are biased low in both methods, except for the two lobes near the high-velocity wings, where the new method performs better. The line width maps (Figure 4) are similar to the intensity except that they are biased

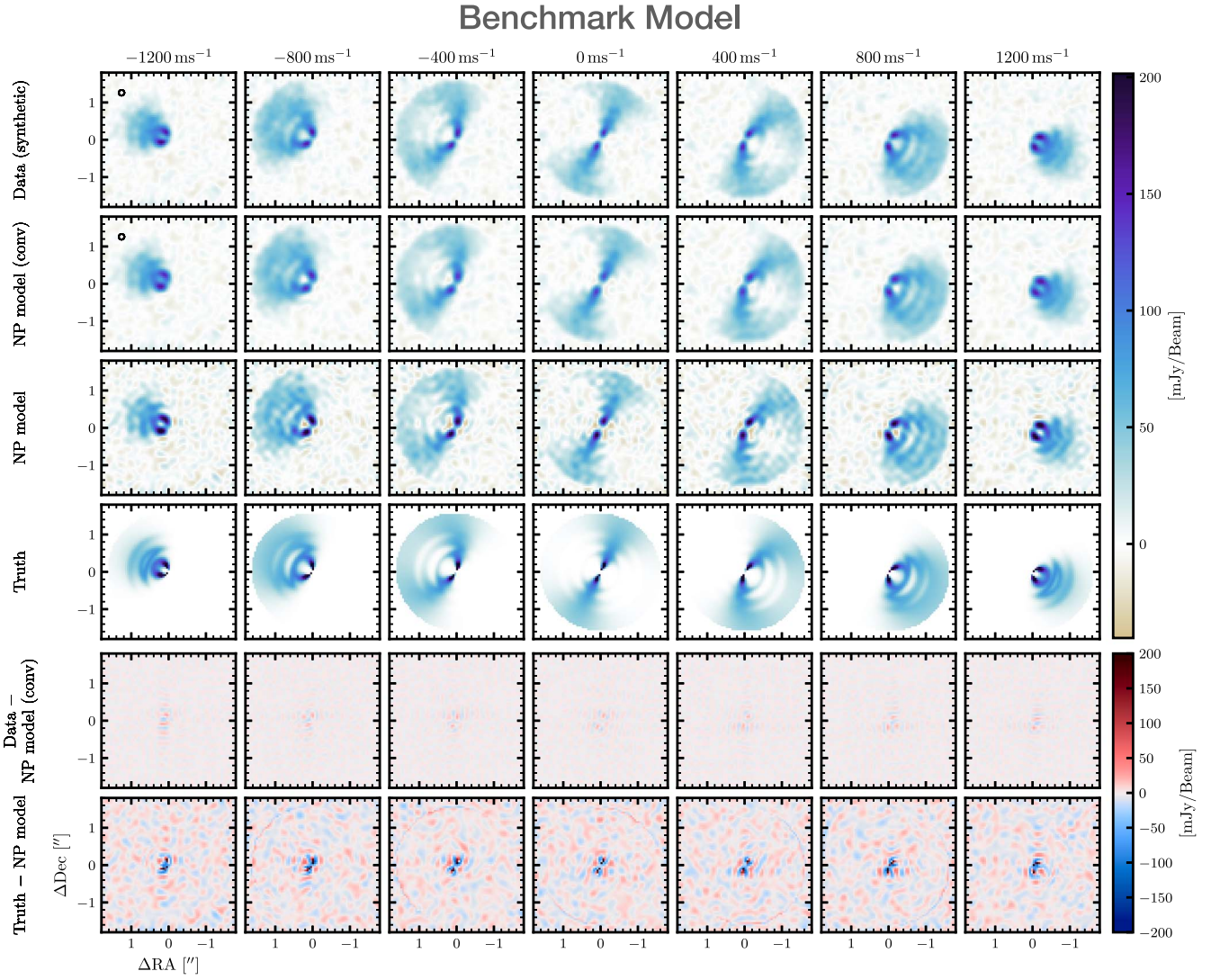


Figure 2. Best-fitting nonparametric (NP) model cube fit to the benchmark model synthetic data, which has a $0''.15$ beam and noise comparable to the data. The top row shows synthetic data. The second row shows the best fit after beam convolution (beam plotted in top left). The third row shows the best fit without beam convolution. The fourth row shows the ground truth for comparison. We see that the model can resolve intensity gradients smeared by the beam. The fifth row shows the residual between the synthetic data and the convolved model. The sixth row shows the residuals between the model and ground truth. The residuals show that the model reconstructs the true intensity distribution up to the noise, except for in the central two beams or so.

high in the conventional method and the bias is greater, around 25%. The new method is again unbiased apart for the same inner region.

4. Results

Figure 5 shows the best-fitting nonparametric model for HD 135344B compared with the data. As in the benchmark, the residuals are all on the scale of the noise or less, and again the most significant difference is found toward the center. The model prior to convolution contains steeper intensity gradients near the obvious arc-like substructures found in the channels, although the noise is increased.

4.1. Moment Maps

The probabilistic v_0 , I_0 , and w_0 maps for HD 135344B, CQ Tau, and MWC 758 are presented in Figures 6 and 7, and we compare our method to the conventional approach. The difference between the two methods is also shown, as well as the 1σ uncertainty for each of the maps. There are common

features shared across the results from the three sources. As in the benchmark, the new method finds greater peak intensities and reduced line widths throughout the disk in each case. The new method also finds greater velocity perturbation amplitudes associated with disk substructures, particularly for HD 135344B. This difference in recovered perturbation size also leads to coherent spiral structures in the panels showing the difference between the v_0 maps for each method. These disks are all known to contain spirals (Y. Boehler et al. 2018; S. Casassus et al. 2021; L. Wölfer et al. 2021; A. F. Izquierdo et al. 2025), and so the model’s ability to recover them, in combination with the benchmark, provides reassurance that it is accurately reflecting the data rather than introducing artificial features.

The maps for HD 135344B (Figure 6, top) show clear signs of large-scale spiral structures spanning across the whole face of the disk in both the velocities and line widths. Our method resolves both of these better than the conventional one, finding velocity perturbations along the spiral 50–100 m s^{-1} larger. The plotted contours on the v_0 maps (top row) show how this

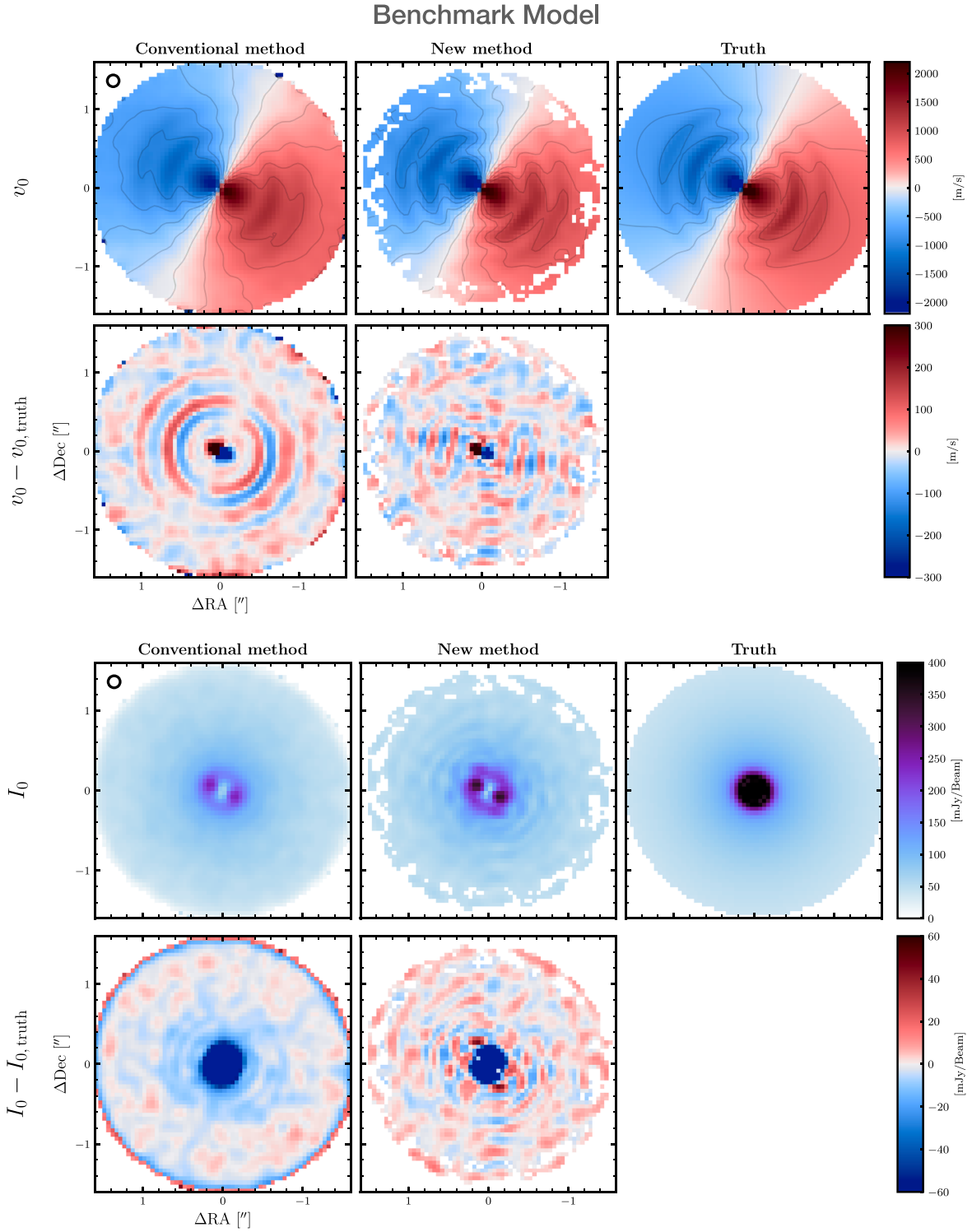
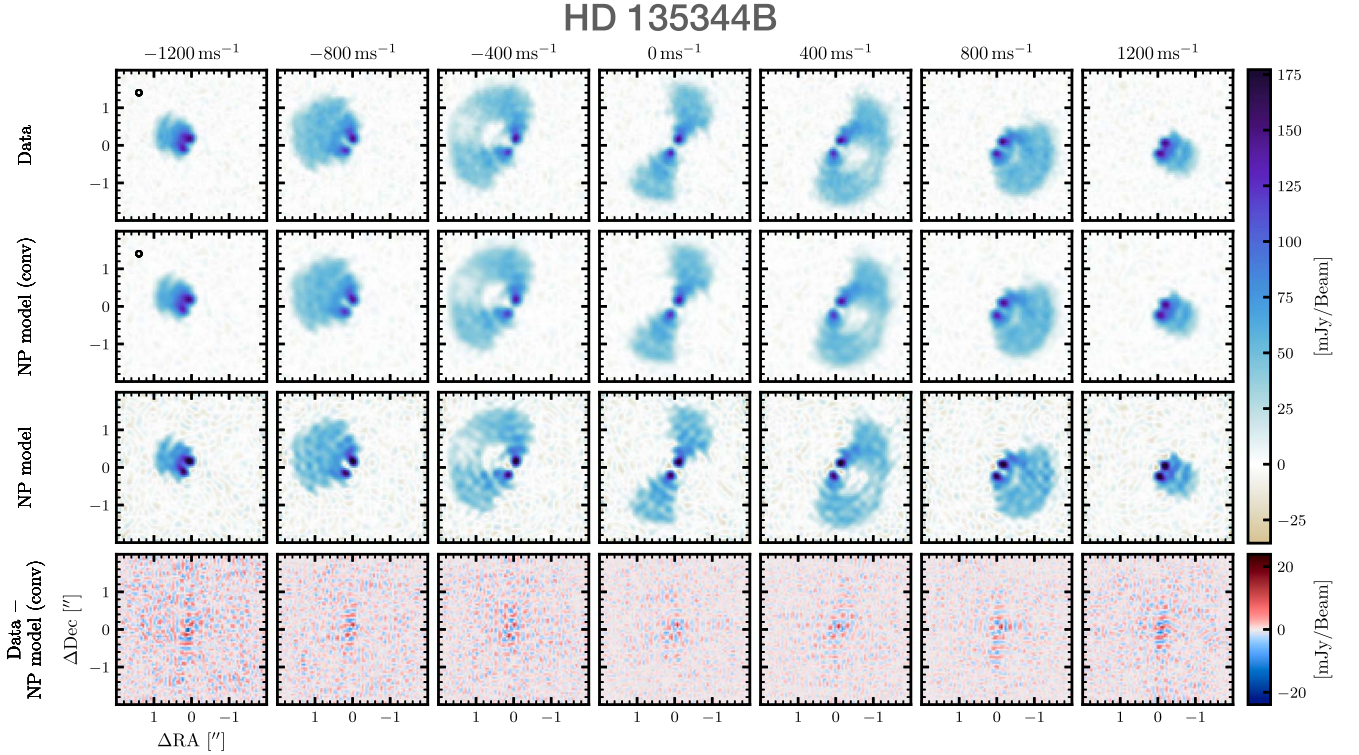
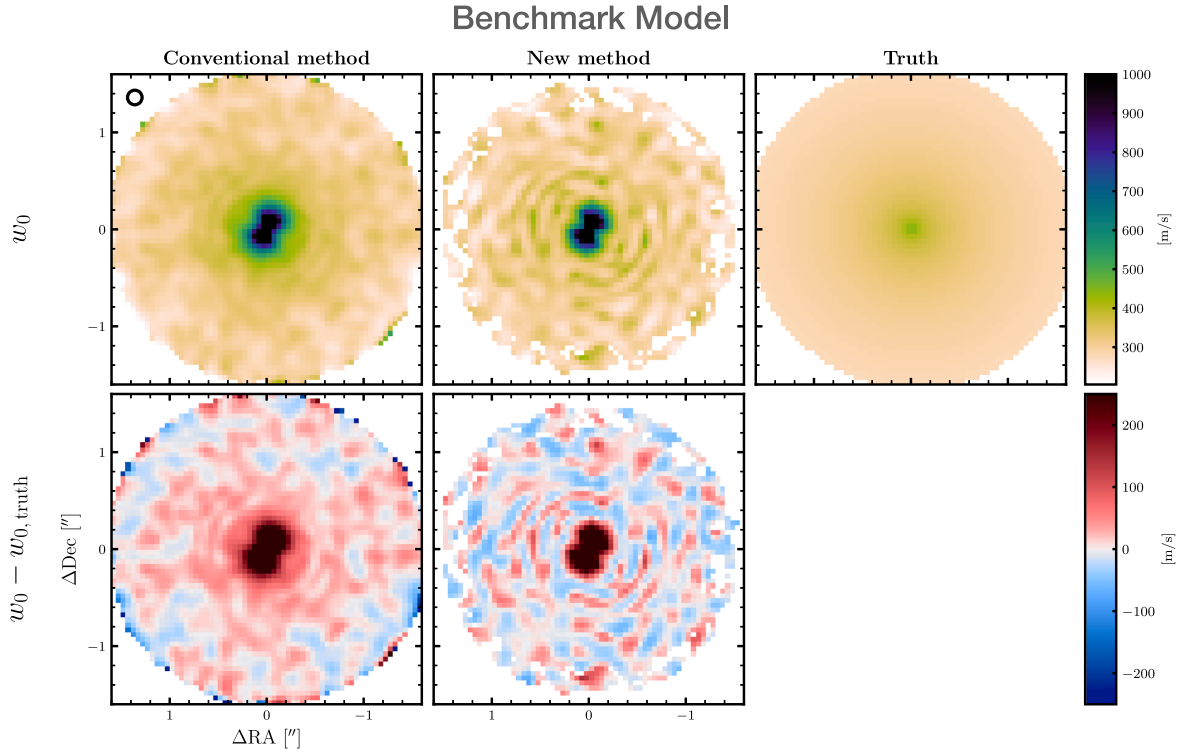


Figure 3. Top: peak velocity maps v_0 extracted from the benchmark model synthetic data, using both the conventional method (left; see Section 2.5) and our method (right). The right-most panel shows the true benchmark model line-of-sight velocities $v_{0, \text{truth}}$. The v_0 maps also show isovelocity contours spaced by 200 m s^{-1} , and the beam is shown in the top left panel. The second row shows the difference between the recovered v_0 maps and the truth. Our method debiases the v_0 map and recovers the full amplitude of the substructures. Bottom: same as above except showing recovered peak intensity maps I_0 .

difference manifests as much larger and sharper wiggles in the isovelocity curves, similar to that in the benchmark. Interestingly, the line widths (bottom row) *along* the spiral are found to be larger than with the conventional method, even though the general trend is lower. The combined effect of smaller line widths in most of the disk and larger line widths along the spiral results in increased contrast of the spiral in the w_0 map.

The I_0 map (middle row), on the other hand, does not differ much between the methods in terms of substructures, and the difference plot shows a general trend toward higher intensities in the new method.

The new method also recovers stronger velocity perturbations in CQ Tau (Figure 6, bottom), with a difference along the large one-armed spiral in the top left of $50\text{--}150 \text{ m s}^{-1}$. Unlike



for HD 135344B, the w_0 map (bottom row) does not show significant substructure, but the I_0 map (middle row) does show hints of the spiral found in v_0 (top row) and is more obvious in the new method result.

MWC 758, unlike the other two sources, shows clear signs of substructure in all three maps (Figure 7, bottom). Our

method again finds larger velocity perturbations (top row) associated with the spiral structure, in the $50\text{--}150\text{ m s}^{-1}$ range, and the contour wiggles are sharper. Both the w_0 and I_0 maps (middle and bottom rows) show increased contrast of the spiral.

The 1σ uncertainty on the v_0 maps correlates with I_0 in all three sources, which is to be expected as the line center

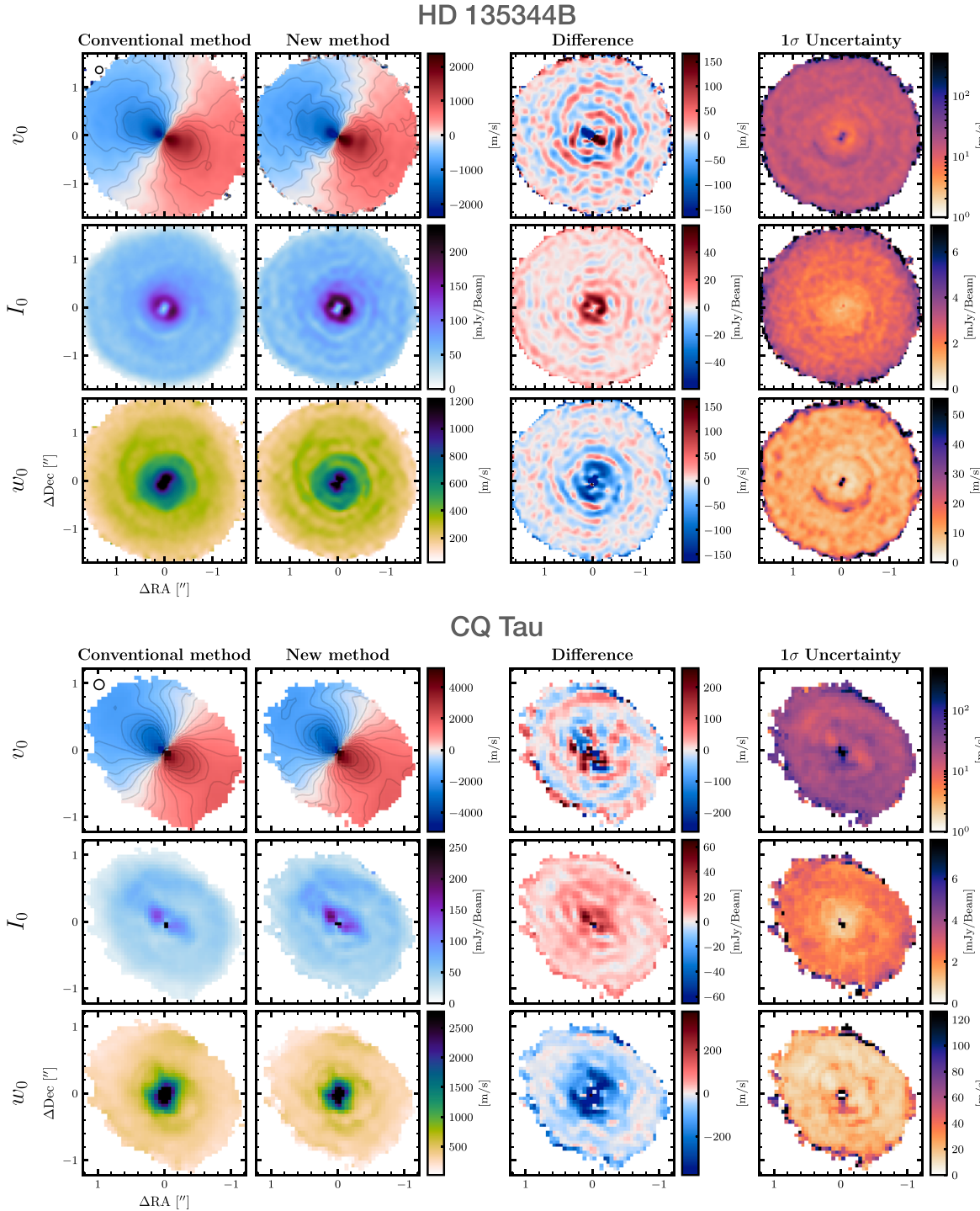


Figure 6. Peak velocity, peak intensity, and line width maps for HD 135344B (top) and CQ Tau (bottom). The left column shows the results using the conventional method. The second column shows the results using our new method. The third column shows the difference of these. The fourth column shows the 1σ uncertainty on the results from the new method. As in the benchmark, the new method better resolves the velocity perturbations associated with substructures.

measurement precision is ultimately set by the signal-to-noise ratio. HD 135344B shows the highest precision, with an uncertainty of $5\text{--}10\text{ m s}^{-1}$ across most of the disk (note the same color map scaling for the v_0 uncertainties in all figures). Both CQ Tau and MWC 758 have uncertainties in the range $10\text{--}30\text{ m s}^{-1}$. This demonstrates that, despite the spatially correlated noise, we can still constrain the line center with a precision comparable to or better than the native spectral resolution of 26 m s^{-1} (R. A. Loomis et al. 2025), which

ultimately sets the fundamental velocity resolution in the absence of additional information about the line profile.

4.2. DISCMINER

In Figure 8, we show the best-fitting nonparametric model to J1842 for one channel compared to the data in the top row. The residuals are mostly noise. The bottom row shows three copies of the same channel from different sample cubes drawn from

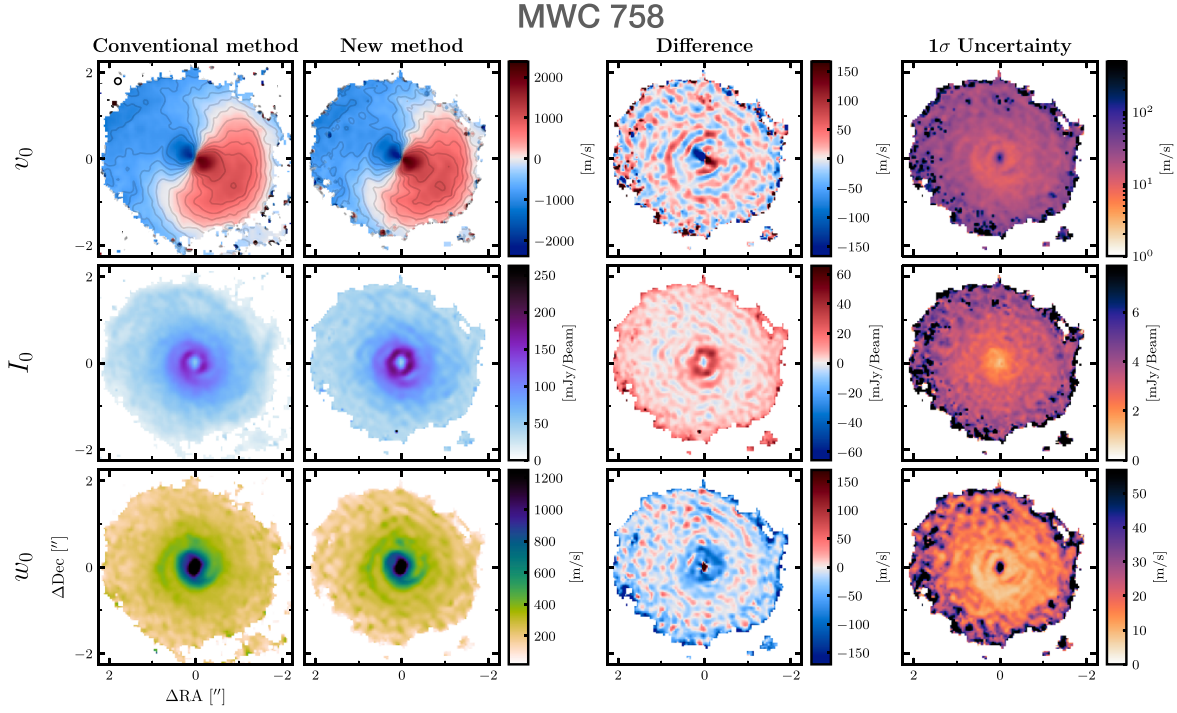


Figure 7. The same as Figure 6 except for MWC 758.

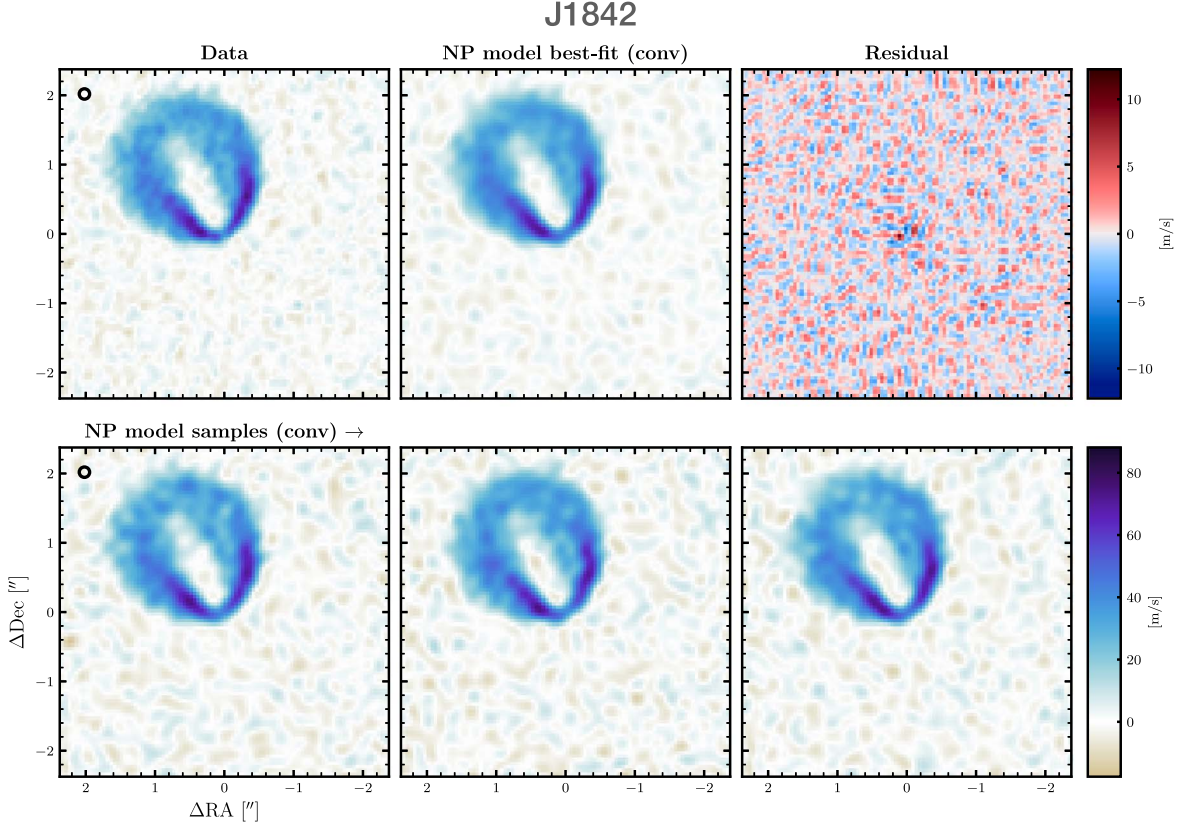


Figure 8. Single channel for the best-fitting nonparametric (NP) model and posterior samples for J1842. The top left panel shows the data and beam. The top middle panel shows the best-fitting model convolved with the beam. The top right panel shows the residual between the data and best fit. The bottom row shows three posterior model channel samples convolved with the beam.

the model posterior and convolved with the beam to demonstrate the typical difference between samples. The large and bright structures are constant across the samples, while the smaller and dimmer features show more variation. Thirty such

samples were used with DISCMINER for the combined approach outlined in Section 2.6.

Figure 9 shows the posterior distributions from the DISCMINER results. We compare the distributions from running

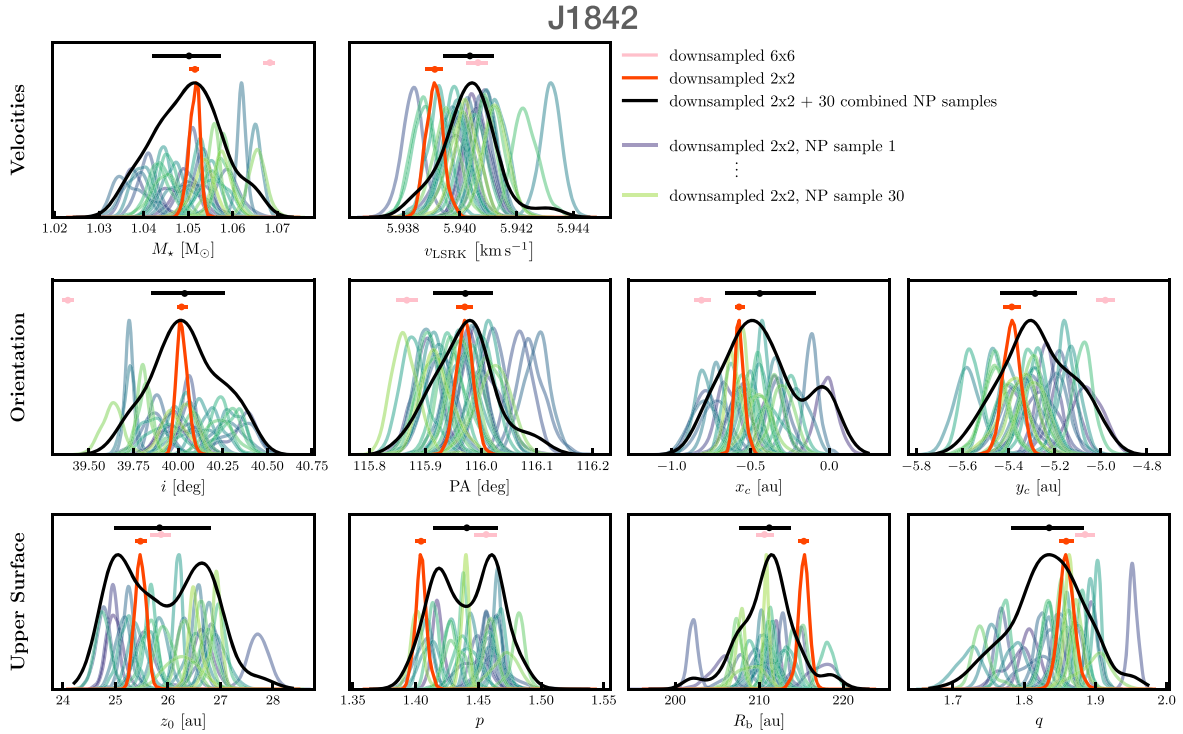


Figure 9. Posterior distributions for a subset of the J1842 DISCMINER model. Posteriors are plotted in each panel, with the corresponding median and 1σ quantiles indicated by the points and lines above the distributions. Fits from individual cube samples are shown in the blue–green color map, and the combined distribution is shown in black. Red shows the posterior from fitting the data directly. Pink is taken from the production exoALMA model. The y-scale of the red and black curves has been scaled for easier comparison. The DISCMINER model parameters shown are stellar mass M_* , systemic velocity v_{LSRK} , inclination i , position angle PA, model origin (x_c, y_c) , reference height z_0 , height power-law index p , taper radius R_b , and taper index q (see A. F. Izquierdo et al. 2025).

DISCMINER on the data directly (red) with the those from running on individual cube samples (blue–green color map) as well as the combined distribution from all cube samples (black). As outlined in Section 2.6, this combined approach allows us to approximate the results that DISCMINER would find, were it to account for the spatially correlated noise.

We find that for all the parameters, the uncertainties are increased by approximately an order of magnitude. The exception is the systemic velocity v_{LSRK} , where the increase is only a factor of a few.

The posterior medians are not different between the data and combined approach to a statistically significant level, indicating that the nonparametric model is not causing a bias. A few values are slightly offset, namely, v_{LSRK} and the upper surface parameters p and R_b . The differences in median surface parameters could be related to the fact that the parameters cause mostly spatially small changes to the shape of the channels and so are potentially affected by accounting for the correlated noise.

Figure 9 also shows the posterior median and 1σ quantiles for the production DISCMINER model fit to the data with additional downsampling. These estimates are mostly compatible with the others, except for M_* and i , which are offset to a significant level even accounting for the increased uncertainty from the correlated noise. This could be caused by the extra downsampling inducing spurious smoothing, biasing the inclination low and the mass high to adjust. While this bias is only on the order of 1%, it is statistically significant and important to consider if the uncertainties from the analysis are to be used.

Since the combined posterior distributions shown are only from 30 cube samples, their overall shape is likely not

accurately reconstructed, and so their non-Gaussian appearance in Figure 9 should not be taken too seriously, including the apparent bimodality in z_0 and p . However, we verified that the estimates of the medians and quantiles become stable at around 10–20 cube samples.

5. Discussion

We have demonstrated that our method for generating probabilistic moment maps is capable of debiasing line-of-sight velocity measurements associated with partially resolved substructures. The correction is important for any studies looking to measure planet masses (F. Bollati et al. 2021), disk masses (B. Veronesi et al. 2021; G. Lodato et al. 2023; C. Longarini et al. 2025; P. Martire et al. 2024; J. Speedie et al. 2024), or vertical shear instability strength (M. Barraza-Alfaro et al. 2021; M. Barraza-Alfaro et al. 2025) directly from the strength of velocity perturbations, as the bias would result in the underestimation of those quantities. For example, measurements of disk mass using the so-called gravitational instability “wiggle” (C. Longarini et al. 2021; J. P. Terry et al. 2022) rely on comparing the deflection of the isovelocity curve near systemic to an analytic model. As we have shown, these wiggles are systematically underestimated by conventional methods but (at least partially) corrected for by our approach. Analyses within exoALMA that rely on measuring velocity perturbations, for example, the measurement of rotation curve deviations (J. Stadler et al. 2025), may be similarly affected, although an in-depth study is beyond the scope of this paper. Our method is less useful in the limits where the substructures are either not resolved or well resolved. As seen in C. Pinte et al. (2025), the sample contains many partially resolved

substructures. The model is also likely to perform significantly worse for low signal-to-noise observations, although we did not test this.

The debiasing effect may also be important for measurements that rely on intensity, such as temperature and pressure, both of which exhibit substructures in the exoALMA sample (M. Galloway-Sprietsma et al. 2025; J. Stadler et al. 2025). Measures of turbulence may also be affected by the reduced line widths recovered with our method (R. Teague et al. 2016), although forward models that fit to visibilities are not subject to bias from beam smearing (K. M. Flaherty et al. 2015; K. Flaherty et al. 2020). The additional contrast we found in the line widths along the substructures may be useful for searches of embedded planets through line broadening (A. F. Izquierdo et al. 2025).

Additionally, our method provides uncertainties that take into account the spatially correlated nature of the noise. Since many samples of each map are created, it is straightforward to propagate this uncertainty downstream into data products. This could be used to quantify the correlation in adjacent data points along extracted rotation curves, as existing approaches are not able to take this into account (C. Longarini et al. 2025).

As alluded to in Section 2.4, the method is applicable to any measurement derived from the data. For example, while we assumed Gaussian line profiles for the creation of moment maps, there is nothing about the method that requires this choice. Recent improvements that more realistically characterize the line profile using higher-order functions or multiple components (R. Teague & D. Foreman-Mackey 2018; A. F. Izquierdo et al. 2025) could be used in place of the Gaussians assumed here. The only stipulation is that it must be computationally feasible to calculate the data product over many sample cubes.

Moreover, the nonparametric model itself makes few assumptions about the form of the data. While we used the fiducial exoALMA ^{12}CO images here, in principle, the data may have any channel spacing or beam size and shape. At full spectral resolution, adjacent channels will be correlated (R. A. Loomis et al. 2018), which we do not account for. More broadly, the model may be applied to any image subject to a point-spread function and noise, provided that the point-spread function is known.

By combining our approach with DISCMINER, we estimated how the accuracy of the inferred model is affected when accounting for spatially correlated noise. For J1842, the uncertainties are increased by an order of magnitude. Despite this, M_* and i are still constrained to within approximately 1%. Additionally, we found that it only takes around 10–20 cube samples for the uncertainty estimates to stabilize. The computational cost for this calculation is greater than that involved in fitting DISCMINER to the data but is still achievable in a few days on a 64 core node. The cost is eased by the fact that all runs after the first can be initialized near the typical set, as the difference between the cube samples is small. For studies concerned with reliable uncertainties on the outputs of DISCMINER, this approach is viable. This uncertainty is purely statistical, however, and if the data are more complicated than assumed in the model, there will be systematic uncertainty that is unaccounted for. This is the case in exoALMA, as the DISCMINER model assumes Keplerian rotation when in reality there are deviations due to self-gravity, pressure support, and substructures (A. F. Izquierdo et al. 2025; C. Longarini et al. 2025;

J. Stadler et al. 2025). The combined approach could also be extended to calculate uncertainties on data products derived using DISCMINER, as each cube sample has its own associated inferred disk model. This would allow for probabilistic 2D residuals, which would aid in probing the robustness of features like Doppler flips (S. Pérez et al. 2018, 2020; C. Pinte et al. 2023), spirals (R. Teague et al. 2021, 2022; J. Calcino et al. 2022; A. F. Izquierdo et al. 2023, 2025), and vortices (P. Huang et al. 2018b; Y. Boehler et al. 2021; L. Wölfer et al. 2025) while accounting for both the correlated noise in the data and the uncertainty in the subtracted disk model.

The model presented in this paper has its own unique advantages, but it is similar to existing methods in the literature. The model itself is of the true intensity distribution in each channel, so it is similar to CLEAN (J. A. Högbom 1974) or regularized maximal likelihood (RML) imaging (K. Akiyama et al. 2017; Event Horizon Telescope Collaboration et al. 2019; B. Zawadzki et al. 2023; B. Zawadzki et al. 2025). Unlike those methods, our model fits to the already-imaged data and not the raw visibilities. We argue that our method therefore constitutes image *analysis* and not image *creation* like in CLEAN or RML, although we admit this distinction is not sharply defined. RML is likely to provide higher accuracy, as it utilizes more information. The benefit provided by our model is that it is linear and so the posterior distribution has an analytic form that is cheap to sample from. While in principle it may be possible to sample from the RML posterior, it may prove intractable, even with samplers like Hamiltonian Monte Carlo that scale well to high dimensions (S. Duane et al. 1987; R. Neal 2011).

FRANK (J. Jennings et al. 2020) fits an axisymmetric model to visibilities directly. Our methods share similarities in that both construct a model from a Fourier series (FRANK uses a generalized Fourier series) and regularization with a GP prior. The most practical difference between our approaches is that our model is of the 2D intensity distribution while FRANK fits for the radial intensity profile. FRANK also operates under the empirical Bayes paradigm, where the covariance structure of the GP is nonparametric, whereas our regularization is equivalent to an exponential covariance kernel.

N. Dia et al. (2023) presented a method to perform image reconstruction in the framework of Bayesian inference, thus providing a posterior distribution over the images. Their method uses score-based priors calculated from a neural network trained on many images of galaxies. This approach is promising but hinges on the choice of training set for the prior. For their test case, continuum images of HD 143006 and AS 209 from the DSHARP program (S. M. Andrews et al. 2018; J. Huang et al. 2018a), the model produced biased fits due to the mismatch in the prior. Variational Bayesian methods such as flows using deep learning have also been shown to be capable of providing tractable sampling from the image posterior (H. Sun & K. L. Bouman 2021), although these methods are formally approximate.

5.1. Caveats

We assumed that the beam is Gaussian. This is true for the CLEAN beam, which is obtained by matching a Gaussian to the synthesized beam’s central component. The synthesized beam contains side lobes, meaning that in the final image, the noise is not correlated exactly according to the CLEAN beam (T. Tsukui et al. 2022). We suspect that for the exoALMA data, the impact on the analysis would be minor, due to the large

number of baselines that ALMA has. Parametric analyses on ALMA line emission data have found that switching from fitting in the image plane to complex visibilities results in negligible difference (K. M. Flaherty et al. 2018). It is unclear whether this generalizes to our model, however, due to the increased flexibility as compared to parametric approaches. The impact of this on the model could be tested by processing the benchmark in a more realistic way with finite sampling of the uv -plane, but we leave this for future work.

The unconvolved model channels show boundary artifacts in the outermost couple of pixels. This is caused by representing the beam convolution with matrix multiplication, as this implicitly assumes that the pixel values are 0 outside the image. This problem is minimized by ensuring that the images are cropped such that a few signal-free pixels are included along the edges.

With the implementation described here, the complexity of the model scales $\mathcal{O}(n^2p + p^3)$, since we fit with numerical least squares (see Appendix B). This makes it prohibitive to use more pixels than around 100×100 . The memory required to store A also approaches 1 TB at around 125×125 pixels. Conjugate gradient descent (M. R. Hestenes & E. Stiefel 1952) would provide a dramatic speed-up in solve time (e.g., J. R. Shewchuk 1994) but does not solve the memory issue. Methods that avoid the construction of A , instead representing the system using linear operators (D. C.-L. Fong & M. Saunders 2011; M. Ravasi & I. Vasconcelos 2020), are a potential avenue for massive speed increases and memory requirement decreases.

As already touched on, our model fits to the CLEAN image and not the raw visibilities. This is a drawback, as information is being discarded. The model is also subject to biases or correlations induced through the CLEANing process. Our approach hinges on the GP prior for the image itself, where the Fourier series is a computational trick more than a model of visibility space. It is not immediately clear how this approach could be modified to fit visibilities instead. A GP prior could be used for the visibilities, although a kernel that encourages smoothness would not be appropriate. This kind of regularization would be weaker than that used in successful approaches like RML, and so it is unclear whether or not it would be viable.

6. Conclusions

We present a new method for generating probabilistic data products from spectral line data of protoplanetary disks and apply it to a subset of the exoALMA sample. Our main findings are summarized as follows.

1. We model the data with a flexible linear model of the image intensities prior to beam convolution, assuming spatially correlated noise. The model is composed of a Fourier basis, where modes with higher frequencies are increasingly suppressed.
2. Model linearity results in an analytic posterior. This allows for the fast generation of sample cubes, which can be used to estimate the posterior distribution for any data product, taking into account the correlated noise in the data.
3. As an example, we presented probabilistic peak velocity, peak intensity, and line width maps of HD 135344B, CQ Tau, and MWC 758. We show that our method more accurately reconstructs line-of-sight velocities associated

with substructures than with conventional methods, which underestimate localized perturbations due to beam smearing. Our method recovers perturbation amplitudes $50\text{--}150\text{ m s}^{-1}$ larger than conventional methods. This debiasing allows more accurate estimates of planet masses, disk masses, vertical shear instability strength, and other analyses that rely on measuring deviations from Keplerian rotation.

4. We also combined our approach with DISCMINER, using J1842 as a case study. We found that when taking into account the spatially correlated noise in the data, the uncertainties on the DISCMINER model parameters increase by an order of magnitude.

The following data products will be publicly available:

1. cube samples for our fits to HD 135344B, CQ Tau, MWC 758, and J1842 and
2. probabilistic peak velocity, peak intensity, and line width maps for HD 135344B, CQ Tau, and MWC 758, including the best fit, uncertainties, and individual samples.

Acknowledgments

This paper makes use of the following ALMA data: ADS/JAO.ALMA#2021.1.01123.L. ALMA is a partnership of ESO (representing its member states), NSF (USA) and NINS (Japan), together with NRC (Canada), MOST and ASIAA (Taiwan), and KASI (Republic of Korea), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by ESO, AUI/NRAO and NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc. We thank the North American ALMA Science Center (NAASC) for their generous support including providing computing facilities and financial support for student attendance at workshops and publications. This work was performed in part on the OzSTAR national facility at Swinburne University of Technology. The OzSTAR program receives funding in part from the Astronomy National Collaborative Research Infrastructure Strategy (NCRIS) allocation provided by the Australian Government and from the Victorian Higher Education State Investment Fund (VHESIF) provided by the Victorian Government. T.H., C.H., and I.H. are supported by Australian Government Research Training Program (RTP) scholarships. The Flatiron Institute is a division of the Simons Foundation. J.B. acknowledges support from NASA XRP grant No. 80NSSC23K1312. M.B., D.F., and J.S. have received funding from the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation program (PROTOPLANETS, grant agreement No. 101002188). Computations by J.S. have been performed on the “Mesocentre SIGAMM” machine, hosted by Observatoire de la Côte d’Azur. P.C. acknowledges support by the Italian Ministero dell’Istruzione, Università e Ricerca through the grant Progetti Premiali 2012—iALMA (CUP C52I13000140001) and by the ANID BASAL project FB210003. S.F. is funded by the European Union (ERC, UNVEIL, 101076613) and acknowledges the financial contribution from PRIN-MUR 2022YP5ACE. M.F. is supported by a grant-in-aid from the Japan Society for the Promotion of Science (KAKENHI; No. JP22H01274). J.D.I. acknowledges support from an STFC Ernest Rutherford Fellowship (ST/W004119/1) and a University Academic Fellowship from the University of Leeds.

Support for A.F.I. was provided by NASA through the NASA Hubble Fellowship grant No. HST-HF2-51532.001-A awarded by the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., for NASA, under contract NAS5-26555. C.L. has received funding from the European Union’s Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 823823 (DUSTBUSTERS) and by the UK Science and Technology Research Council (STFC) via the consolidated grant ST/W000997/1. C.P. and D.P. acknowledge Australian Research Council funding via FT170100040, DP18010423, DP220103767, and DP240103290. A.C. acknowledges Australian Research Council funding via DP210100018. G.R. acknowledges funding from the Fondazione Cariplo, grant No. 2022-1217, and the European Research Council (ERC) under the European Union’s Horizon Europe Research & Innovation Program under grant agreement No. 101039651 (DiscEvol). H.-W.Y. acknowledges support from the National Science and Technology Council (NSTC) in Taiwan through grant NSTC 113-2112-M-001-035- and from the Academia Sinica Career Development Award (AS-CDA-111-M03). G.W.F. acknowledges support from the European Research Council (ERC) under the European Union Horizon 2020 research and innovation program (grant agreement No. 815559, MHDiscs). G.W.F. was granted access to the HPC resources of IDRIS under the allocation A0120402231 made by GENCI. Support for B.Z. was provided by The Brinson Foundation. Views and opinions expressed by ERC-funded scientists are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Software: Astropy (Astropy Collaboration et al. 2022), NumPy (C. R. Harris et al. 2020), SciPy (P. Virtanen et al. 2020), matplotlib (J. D. Hunter 2007), cmasher (E. van der Velden 2020).

Appendix A Impact of Hyperparameter Selection

As the hyperparameters λ and s appear directly in the expression for the model posterior mean (Equation (16)) and covariances (Equation (17)), the choice of their values will impact both the predictions and uncertainties in our results. Increasing either λ or s in general leads to spatially smoother predictions and decreased variance in the model. Setting λ and s to be too large will result in models that smear out the substructures we are interested in, while setting them too small will result in model images with too much high-frequency noise.

Ideally, one would set the values of λ and s using the data themselves. This can be achieved with cross-validation (CV), where the hyperparameters are tuned iteratively by partitioning the data into testing and training sets and assessing the predictive performance on the held-out data (e.g., M. W. Browne 2000). For our method, varying the hyperparameters does affect the model predictions, but the convolved model predictions are altered only negligibly, since the regularization mostly impacts modes of higher frequencies. Since the CV error metric would be between the data and the

convolved model, this avenue does not provide a reliable way to ensure that the model prior to convolution is performant.

Alternatively, we could forward model a set of realistic known images to create synthetic data that could then be used as a means of calibrating the prior and selecting hyperparameter values. While an analysis of this type would strengthen confidence in the hyperparameter choice, a full implementation is out of scope for the immediate focus of this work, in which our priority is to introduce the core methodology. Nonetheless, as demonstrated in the benchmark (Section 3), our chosen hyperparameters yield plausible reconstructions of the intensity gradients around substructures leading to more accurate inferences when compared with the ground truth.

We should also be concerned with the impact of the hyperparameter choices on the uncertainties in our results. To assess this, we calculated Σ for a large range of possible λ values for the fit to the HD 135344B data. We varied only λ and kept $s = 0.1$ fixed because it has been found that for this kernel, only the product $\lambda^2\sqrt{s}$ is actually identifiable from the data in general (H. Zhang 2004; A. Gelman et al. 2017). We then found the marginal standard deviation on the model predictions for each pixel before and after convolution as

$$\sigma(\mathbf{I}) = \sqrt{\text{diag}(\mathbf{A}_{xy} \Sigma \mathbf{A}_{xy}^T)}, \quad (\text{A1})$$

$$\sigma(\mathbf{I}_{\text{conv}}) = \sqrt{\text{diag}(\mathbf{A} \Sigma \mathbf{A}^T)}, \quad (\text{A2})$$

where diag extracts the diagonal of a matrix as a vector. We note that the calculated variance is almost identical in each pixel for a given λ . The resultant values are shown in Figure 10. We see the model prediction uncertainty decreases approximately linearly with $\log \lambda$, while the convolved model uncertainty is essentially flat. At $\lambda \approx 10^4$, there is a turnoff from this behavior where the model uncertainty decreases more rapidly and the convolved model uncertainty begins to decrease as well. This difference is because the regularization acts in the model to suppress high-frequency modes: as the prior becomes weaker, the model can contain more high-frequency noise, but the convolved model is basically unchanged, as the high-frequency information does not survive the convolution. At very large λ , shaded red in the figure, the regularization is strong enough to suppress modes of lower frequencies such that the variance in the convolved model predictions is affected. This corresponds to a prior so strong that the convolved model predictions will be smoother than the beam, so we label this as overregularized.

The green shaded area in Figure 10 corresponds to the level of regularization we used in our results, $\lambda = 10^{-1}$. The downstream data products presented in Section 4.1 depend on per-pixel quantities, and so their uncertainty will also be underestimated if our regularization is too strong. This will likely be a minor effect, as a decrease in regularization of 4 orders of magnitude only increases the uncertainty by a factor of around 30%. For the analysis using DISCMINER presented in Section 4.2, we used samples from the convolved model, and so the results and uncertainties should be unaffected by a different choice of hyperparameters unless overregularized.

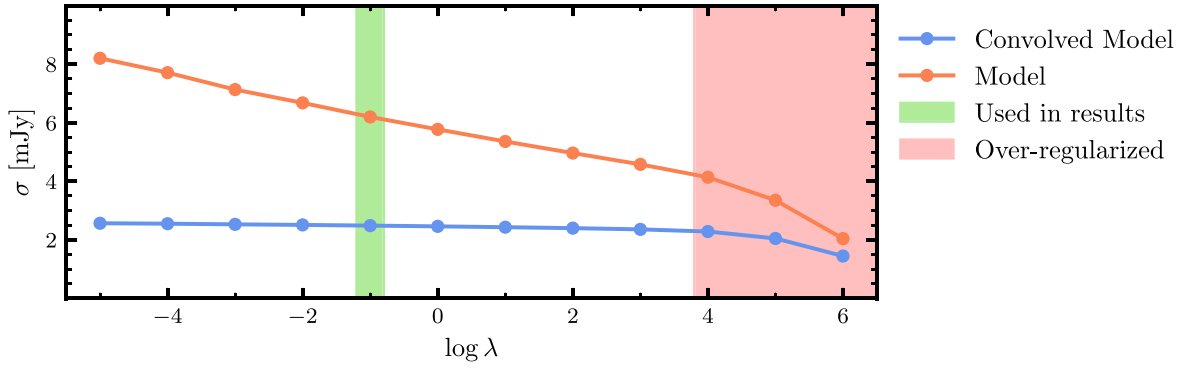


Figure 10. Uncertainty in model prediction as a function of regularization strength λ for the model before (orange) and after (blue) convolution with the beam. The green region indicates the value of λ used in our results. The red region indicates the region where the model is clearly overregularized.

Appendix B Numerical Considerations

In Section 2.2, when outlining our model, we gave the solution for \hat{X} as Equation (16). In practice, we instead solve for the equivalent expression (H. V. Henderson & S. R. Searle 1981) when $\mu = 0$,

$$\hat{X} = \mathbf{A}\mathbf{A}^\top (\mathbf{A}\mathbf{A}^\top + \mathbf{C})^{-1} \mathbf{I}_{\text{data}}, \quad (\text{B1})$$

since it is faster and more numerically stable for the case where $p > n$ (e.g., D. W. Hogg & S. Villar 2021), which we used in this work. It also avoids inverting \mathbf{C} . The solution in Equation (16) is faster and more numerically stable if $n < p$, and we left it in the main text since its connection to the regular GLS solution in Equation (13) is more obvious.

Additionally, the condition number of \mathbf{H} , and so also \mathbf{A} and \mathbf{C} , can be very large due to the small tails of the Gaussian distribution used for the beam kernel h in Equation (6). This problem worsens when the beam is more resolved, i.e., when the number of pixels per beam is increased. We deal with this issue in two ways. First, we downsample the cubes spatially by a factor of 2, which reduces the number of pixels per beam, but not so much so that the correlations between nearby pixels are no longer resolved. This also helps reduce computation cost since it reduces n by a factor of 4. Second, following the advice of D. W. Hogg & S. Villar (2021), we avoid directly calculating the inverse of matrices (with the exception of $\mathbf{\Lambda}^{-1}$, since it is diagonal). We instead solve for products of the inverse using numerical least squares with NUMPY's `linalg.lstsq` function with RCOND set to 10^{-16} to zero out small eigenvalues near machine precision. That is, for some system $\mathbf{A}\mathbf{X} = \mathbf{B}$, we use `linalg.lstsq(A, B)` to find \mathbf{X} instead of ever calculating \mathbf{A}^{-1} . This approach is especially important for

$\mathbf{A}^\top \mathbf{C}^{-1} \mathbf{A} + \mathbf{\Lambda}^{-1}$. Finally, we use an LU decomposition for the final unavoidable direct inversion, yielding Σ .

After calculating \hat{X} and Σ , we need to draw samples from the posterior distribution given by Equation (15), which is a p -dimensional normal distribution. Let \mathbf{Z} be a vector containing p independent draws from the standard normal distribution (zero mean and unit variance), and let \mathbf{B} be any real matrix such that $\Sigma = \mathbf{B}\mathbf{B}^\top$. We can then generate samples \mathbf{X} from the desired distribution using

$$\mathbf{X} = \hat{X} + \mathbf{B}\mathbf{Z}. \quad (\text{B2})$$

In theory, Σ is guaranteed to be symmetric, and it may be positive definite or positive semidefinite. In practice, Σ is often not quite symmetric due to numerical precision, and so we calculated a symmetric version $\Sigma' = 0.5(\Sigma + \Sigma^\top)$, which we use to find \mathbf{B} . If Σ is positive definite, then we use a Cholesky decomposition,

$$\Sigma = \mathbf{L}\mathbf{L}^\top, \quad (\text{B3})$$

where \mathbf{L} is a real lower-triangular matrix. If Σ is only positive semidefinite, we instead use a spectral decomposition,

$$\Sigma = \mathbf{Q}\mathbf{D}\mathbf{Q}^\top = (\mathbf{Q}\mathbf{D}^{1/2})(\mathbf{Q}\mathbf{D}^{1/2})^\top, \quad (\text{B4})$$

where \mathbf{Q} is an orthogonal matrix containing the eigenvectors of Σ as columns and \mathbf{D} is a diagonal matrix whose diagonal elements are the corresponding eigenvalues, which are all nonnegative.

Appendix C Benchmark Model

The radial intensity profile we use is

$$I(r) = \begin{cases} \left[0.35 \exp\left(-\left[\frac{r}{40 \text{ au}}\right]^{0.45}\right) + 0.5 \exp\left(-\frac{1}{2} \left[\frac{r - 7 \text{ au}}{15 \text{ au}}\right]^2\right) \right] \text{Jy beam}^{-1} & r \leq 220 \text{ au}, \\ 0 \text{ Jy beam}^{-1} & \text{elsewhere,} \end{cases} \quad (\text{C1})$$

calculating the uncertainties on \hat{X} since it requires the inverse of \mathbf{C} . We break the solution in Equation (17) into the following steps. We set $\mathbf{\Gamma} = \mathbf{A}^\top \mathbf{C}^{-1}$ giving $\mathbf{\Gamma}\mathbf{C} = \mathbf{A}^\top$, where we can then solve for $\mathbf{\Gamma}$ using the above approach. We then calculate the expression inside the brackets with $\mathbf{\Gamma}\mathbf{A} + \mathbf{\Lambda}^{-1} =$

and the line width profile is

$$w(r) = 0.65 \exp\left(-\left[\frac{r}{500 \text{ au}}\right]^{0.2}\right) \text{km s}^{-1}. \quad (\text{C2})$$

We assume azimuthal Keplerian motions $v(r) = v_\phi(r)$, with three Gaussian ring perturbations placed at different radii,

$$v_\phi(r) = [1 - 0.25 G_{70}(r) + 0.35 G_{90}(r) + 0.25 G_{140}(r)] v_K(r), \quad (\text{C3})$$

$$v_K(r) = \frac{GM_\star}{r}, \quad (\text{C4})$$

$$G_R(r) = \exp\left(-\frac{1}{2} \left[\frac{r - R}{10 \text{ au}} \right]^2\right), \quad (\text{C5})$$

such that the perturbations have amplitudes of -0.25 , 0.35 , and 0.25 Keplerian at radii of 70 , 90 , and 140 au, respectively. The intensity and line width radial profiles were chosen by hand to produce roughly similar channels to HD 135344B, and their exact form should not be treated as a careful fit to the data.

ORCID iDs

Thomas Hilder  <https://orcid.org/0000-0001-7641-5235>
 Andrew R. Casey  <https://orcid.org/0000-0003-0174-0564>
 Daniel J. Price  <https://orcid.org/0000-0002-4716-4235>
 Christophe Pinte  <https://orcid.org/0000-0001-5907-5179>
 Andrés F. Izquierdo  <https://orcid.org/0000-0001-8446-3026>
 Caitlyn Hardiman  <https://orcid.org/0009-0003-7403-9207>
 Jaehan Bae  <https://orcid.org/0000-0001-7258-770X>
 Marcelo Barraza-Alfaro  <https://orcid.org/0000-0001-6378-7873>
 Myriam Benisty  <https://orcid.org/0000-0002-7695-7605>
 Gianni Cataldi  <https://orcid.org/0000-0002-2700-9676>
 Pietro Curone  <https://orcid.org/0000-0003-2045-2154>
 Ian Czekala  <https://orcid.org/0000-0002-1483-8811>
 Stefano Facchini  <https://orcid.org/0000-0003-4689-2684>
 Daniele Fasano  <https://orcid.org/0000-0003-4679-4072>
 Mario Flock  <https://orcid.org/0000-0002-9298-3029>
 Misato Fukagawa  <https://orcid.org/0000-0003-1117-9213>
 Maria Galloway-Sprietsma  <https://orcid.org/0000-0002-5503-5476>
 Himanshi Garg  <https://orcid.org/0000-0002-5910-4598>
 Cassandra Hall  <https://orcid.org/0000-0002-8138-0425>
 Iain Hammond  <https://orcid.org/0000-0003-1502-4315>
 Jane Huang  <https://orcid.org/0000-0001-6947-6072>
 John D. Ilee  <https://orcid.org/0000-0003-1008-1142>
 Kazuhiro Kanagawa  <https://orcid.org/0000-0001-7235-2417>
 Geoffroy Lesur  <https://orcid.org/0000-0002-8896-9435>
 Cristiano Longarini  <https://orcid.org/0000-0003-4663-0318>
 Ryan Loomis  <https://orcid.org/0000-0002-8932-1219>
 Ryuta Orihara  <https://orcid.org/0000-0003-4039-8933>
 Giovanni Rosotti  <https://orcid.org/0000-0003-4853-5736>
 Jochen Stadler  <https://orcid.org/0000-0002-0491-143X>
 Richard Teague  <https://orcid.org/0000-0003-1534-5186>
 Hsi-Wei Yen  <https://orcid.org/0000-0003-1412-893X>
 Gaylor Wafflard  <https://orcid.org/0000-0002-3468-9577>
 Andrew J. Winter  <https://orcid.org/0000-0002-7501-9801>
 Lisa Wölfer  <https://orcid.org/0000-0002-7212-2416>
 Tomohiro C. Yoshida  <https://orcid.org/0000-0001-8002-8473>
 Brianna Zawadzki  <https://orcid.org/0000-0001-9319-1296>

References

- Akiyama, K., Kuramochi, K., Ikeda, S., et al. 2017, *ApJ*, **838**, 1
 Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, *ApJL*, **869**, L41
 Andrews, S. M., Teague, R., Wirth, C. P., Huang, J., & Zhu, Z. 2024, *ApJ*, **970**, 153
 Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., et al. 2022, *ApJ*, **935**, 167
 Bah, B., & Ward, R. 2016, *ITSP*, **64**, 3145
 Barraza-Alfaro, M., Flock, M., Béthune, W., et al. 2025, *ApJL*, **984**, L21
 Barraza-Alfaro, M., Flock, M., Marino, S., & Pérez, S. 2021, *A&A*, **653**, A113
 Bernardo, J., & Smith, A. 2009, *Bayesian Theory* (New York: Wiley)
 Boehler, Y., Ménard, F., Robert, C. M. T., et al. 2021, *A&A*, **650**, A59
 Boehler, Y., Ricci, L., Weaver, E., et al. 2018, *ApJ*, **853**, 162
 Bollati, F., Lodato, G., Price, D. J., & Pinte, C. 2021, *MNRAS*, **504**, 5444
 Browne, M. W. 2000, *J. Math. Psychol.*, **44**, 108
 Broyden, C. G. 1970, *JApMa*, **6**, 76
 Calcino, J., Hilder, T., Price, D. J., et al. 2022, *ApJ*, **929**, L25
 Casassus, S., Christiaens, V., Cárcamo, M., et al. 2021, *MNRAS*, **507**, 3789
 Cotton, W. D. 1989, in *ASP Conf. Ser. 6, Synthesis Imaging in Radio Astronomy*, ed. R. A. Perley, F. R. Schwab, & A. H. Bridle (San Francisco, CA: ASP), **233**
 Czekala, I., Andrews, S. M., Jensen, E. L. N., et al. 2015, *ApJ*, **806**, 154
 Dia, N., Yantovski-Barth, M. J., Adam, A., et al. 2023, arXiv:2311.18012
 Disk Dynamics Collaboration, Armitage, P. J., Bae, J., et al. 2020, arXiv:2009.04345
 Duane, S., Kennedy, A. D., Pendleton, B. J., & Roweth, D. 1987, *PhLB*, **195**, 216
 Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019, *ApJL*, **875**, L4
 Flaherty, K., Hughes, A. M., Simon, J. B., et al. 2020, *ApJ*, **895**, 109
 Flaherty, K. M., Hughes, A. M., Rosenfeld, K. A., et al. 2015, *ApJ*, **813**, 99
 Flaherty, K. M., Hughes, A. M., Teague, R., et al. 2018, *ApJ*, **856**, 117
 Fletcher, R. 1970, *CompJ*, **13**, 317
 Fong, D. C.-L., & Saunders, M. 2011, *SJSC*, **33**, 2950
 Gaia Collaboration, Vallenari, A., Brown, A. G. A., et al. 2023, *A&A*, **674**, A1
 Galloway-Sprietsma, M., Bae, J., Izquierdo, A., et al. 2025, *ApJL*, **984**, L10
 Galloway-Sprietsma, M., Bae, J., Teague, R., et al. 2023, *ApJ*, **950**, 147
 Garg, H., Pinte, C., Hammond, I., et al. 2022, *MNRAS*, **517**, 5942
 Gelman, A., Carlin, J., Stern, H., et al. 2014, *Bayesian Data Analysis* (3rd ed.; Boca Raton, FL: Chapman & Francis)
 Gelman, A., Simpson, D., & Betancourt, M. 2017, *Entrep*, **19**, 555
 Goldfarb, D. 1970, *MaCom*, **24**, 23
 Greengard, P., Rachh, M., & Barnett, A. 2022, *SIAM/ASA Journal on Uncertainty Quantification*, **13**, 63
 Harris, C. R., Millman, K. J., van der Walt, S. J., et al. 2020, *Natur*, **585**, 357
 Henderson, H. V., & Searle, S. R. 1981, *SIAMR*, **23**, 53
 Hensman, J., Durand, N., & Solin, A. 2018, *JMLR*, **18**, 1, <http://jmlr.org/papers/v18/16-579.html>
 Hestenes, M. R., & Stiefel, E. 1952, *JRNBS*, **49**, 409
 Högbom, J. A. 1974, *A&AS*, **15**, 417
 Hogg, D. W., & Foreman-Mackey, D. 2018, *ApJS*, **236**, 11
 Hogg, D. W., Price-Whelan, A. M., & Leistedt, B. 2020, arXiv:2005.14199
 Hogg, D. W., & Villar, S. 2021, *PASP*, **133**, 093001
 Huang, J., Andrews, S. M., Dullemond, C. P., et al. 2018a, *ApJL*, **869**, L42
 Huang, P., Isella, A., Li, H., Li, S., & Ji, J. 2018b, *ApJ*, **867**, 3
 Hunter, J. D. 2007, *CSE*, **9**, 90
 Izquierdo, A. F., Facchini, S., Rosotti, G. P., van Dishoeck, E. F., & Testi, L. 2022, *ApJ*, **928**, 2
 Izquierdo, A. F., Stadler, J., Bae, J., et al. 2025, *ApJL*, **984**, L8
 Izquierdo, A. F., Testi, L., Facchini, S., Rosotti, G. P., & van Dishoeck, E. F. 2021, *A&A*, **650**, A179
 Izquierdo, A. F., Testi, L., Facchini, S., et al. 2023, *A&A*, **674**, A113
 Jennings, J., Booth, R. A., Tazzari, M., Rosotti, G. P., & Clarke, C. J. 2020, *MNRAS*, **495**, 3209
 Kepler, M., Teague, R., Bae, J., et al. 2019, *A&A*, **625**, A118
 Kurtovic, N. T., & Pinilla, P. 2024, *A&A*, **687**, A188
 Kutner, M., Nachtsheim, C., & Neter, J. 2004, *Applied Linear Regression Models*, Irwin/McGraw-Hill Series in Operations and Decision Sciences (New York: McGraw-Hill Education)
 Liu, D. C., & Nocedal, J. 1989, *MatPr*, **45**, 503
 Lodato, G., Rampinelli, L., Viscardi, E., et al. 2023, *MNRAS*, **518**, 4481
 Longarini, C., Lodato, G., Rosotti, G., et al. 2025, *ApJL*, **984**, L17
 Longarini, C., Lodato, G., Toci, C., et al. 2021, *ApJL*, **920**, L41
 Loomis, R. A., Facchini, S., Benisty, M., et al. 2025, *ApJL*, **984**, L7
 Loomis, R. A., Öberg, K. I., Andrews, S. M., et al. 2018, *AJ*, **155**, 182
 Martire, P., Longarini, C., Lodato, G., et al. 2024, *A&A*, **686**, A9
 Moore, E. H. 1920, *BAMaS*, **26**, 394
 Neal, R. 2011, *Handbook of Markov Chain Monte Carlo* (London: Chapman & Hall), **113**
 Oldenburg, D. W. 1981, *GeoJI*, **65**, 331
 Pegues, J., Czekala, I., Andrews, S. M., et al. 2021, *ApJ*, **908**, 42

- Penrose, R. 1956, [MPCPS](#), **52**, 17
- Pérez, S., Casassus, S., & Benítez-Llambay, P. 2018, [MNRAS](#), **480**, L12
- Pérez, S., Casassus, S., Hales, A., et al. 2020, [ApJL](#), **889**, L24
- Pinte, C., Ilee, J. D., Huang, J., et al. 2025, [ApJL](#), 984, L15
- Pinte, C., Teague, R., Flaherty, K., et al. 2023, in ASP Conf. Ser. 534, Protostars and Planets VII, 534, ed. S. Inutsuka et al. (San Francisco, CA: ASP)
- Rasmussen, C. E., & Williams, C. K. I. 2006, Gaussian Processes for Machine Learning (Cambridge, MA: MIT Press)
- Rauhut, H., & Ward, R. 2016, [Appl. Comput. Harmon. Anal.](#), **40**, 321
- Ravasi, M., & Vasconcelos, I. 2020, [SoftX](#), **11**, 100361
- Rosotti, G. P., Teague, R., Dullemond, C., Booth, R. A., & Clarke, C. J. 2020, [MNRAS](#), **495**, 173
- Shanno, D. F. 1970, [MaCom](#), **24**, 647
- Shewchuk, J. R. 1994, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Tech. rep., USA, Carnegie Mellon Univ.
- Speedie, J., Dong, R., Hall, C., et al. 2024, [Natur](#), **633**, 58
- Stadler, J., Benisty, M., Winter, A. J., et al. 2025, [ApJL](#), 984, L11
- Sun, H., & Bouman, K. L. 2021, in Proc. of the AAAI Conf. on Artificial Intelligence, 35, 2628
- Teague, R., Bae, J., Aikawa, Y., et al. 2021, [ApJS](#), **257**, 18
- Teague, R., Bae, J., Andrews, S. M., et al. 2022, [ApJ](#), **936**, 163
- Teague, R., Bae, J., & Bergin, E. A. 2019, [Natur](#), **574**, 378
- Teague, R., Bae, J., Bergin, E. A., Birnstiel, T., & Foreman-Mackey, D. 2018, [ApJL](#), **860**, L12
- Teague, R., Benisty, M., Facchini, S., et al. 2025, [ApJL](#), 984, L6
- Teague, R., & Foreman-Mackey, D. 2018, [RNAAS](#), **2**, 173
- Teague, R., Guilloteau, S., Semenov, D., et al. 2016, [A&A](#), **592**, A49
- Terry, J. P., Hall, C., Longarini, C., et al. 2022, [MNRAS](#), **510**, 1671
- Tobar, F., Robert, A., & Silva, J. F. 2023, [RSPSA](#), **479**, 20220648
- Tsukui, T., Iguchi, S., Mitsuhashi, I., & Tadaki, K. 2022, [Proc. SPIE](#), **12190**, 121901C
- van der Velden, E. 2020, [JOSS](#), **5**, 2004
- Veronesi, B., Paneque-Carreño, T., Lodato, G., et al. 2021, [ApJL](#), **914**, L27
- Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, [NatMe](#), **17**, 261
- Wiener, N. 1949, Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications (Cambridge, MA: MIT Press)
- Wölfer, L., Barraza-Alfaro, M., Teague, R., et al. 2025, [ApJL](#), 984, L22
- Wölfer, L., Facchini, S., Kurtovic, N. T., et al. 2021, [A&A](#), **648**, A19
- Xie, Y., Chou, H.-H., Rauhut, H., & Ward, R. 2022, [SIMODS](#), **4**, 885
- Zawadzki, B., Czekala, I., Galloway-Sprietsma, M., et al. 2025, [ApJL](#), 984, L14
- Zawadzki, B., Czekala, I., Loomis, R. A., et al. 2023, [PASP](#), **135**, 064503
- Zhang, H. 2004, [JASA](#), **99**, 250