Frankenstein: Protoplanetary disc brightness profile reconstruction at sub-beam resolution with a rapid Gaussian process

Jeff Jennings,[1][*] Richard A. Booth,[1] Marco Tazzari,[1] Giovanni P. Rosotti,[2] and Cathie J. Clarke[1]

[1]*Institute of Astronomy, University of Cambridge, Madingley Road, Cambridge, CB3 0HA, UK*
[2]*Leiden Observatory, University of Leiden, P.O. Box 9500, Leiden, NL-2300 RA, Netherlands*

ABSTRACT

Interferometric observations of the mm dust distribution in protoplanetary discs are now showing a ubiquity of annular gap and ring substructures. Their identification and accurate characterization is critical to probing the physical processes responsible. We present Frankenstein (`frank`), an open source code that recovers axisymmetric disc structures at sub-beam resolution. By fitting the visibilities directly, the model reconstructs a disc's 1D radial brightness profile nonparametrically using a fast ($\lesssim 1$ min) Gaussian process. The code avoids limitations of current methods that obtain the radial brightness profile by either extracting it from the disc image via nonlinear deconvolution at the cost of reduced fit resolution, or by assumptions placed on the functional forms of disc structures to fit the visibilities parametrically. We use mock ALMA observations to quantify the method's intrinsic capability and its performance as a function of baseline-dependent signal-to-noise. Comparing the technique to profile extraction from a CLEAN image, we motivate how our fits accurately recover disc structures at a sub-beam resolution. Demonstrating the model's utility in fitting real high and moderate resolution observations, we conclude by proposing applications to address open questions on protoplanetary disc structure and processes. 🜨 📄

Key words: techniques: interferometric, submillimetre: general, submillimetre: planetary systems, protoplanetary discs, planets and satellites: detection, methods: data analysis

1 PHYSICAL AND METHODOLOGICAL CONTEXTS

High resolution interferometric observations in the sub-mm – mm and cm are now providing a detailed picture of protoplanetary disc structure, offering an opportunity for new insights into not only disc morphologies but also the underlying physical properties. Measurements with the Atacama Large Millimeter Array (ALMA) and the Karl G. Jansky Very Large Array (VLA) are finding a ubiquity of annular 'gaps' and 'rings' in disc emission. At the highest angular resolutions achieved to-date (30 – 65 mas corresponding to $\approx 5 - 10$ AU)[1], these features are seen or hinted at in essentially all bright discs around single stars (e.g., ALMA Part-

nership et al. 2015; Andrews et al. 2016; Clarke et al. 2018; Huang et al. 2018; Kudo et al. 2018; Keppler et al. 2019; Pinte et al. 2019; Pérez et al. 2019). Unbiased disc surveys at moderate resolution support this trend, with annular substructure present in at least 38% (12 / 32) of discs in the Taurus survey at ≈ 120 mas (≈ 15 AU; Long et al. 2018) and hinted at in $\approx 5\%$ (3 / 53) Ophiuchus discs at ≈ 200 mas (≈ 28 AU; Cieza et al. 2019). Intriguingly this suggests that many apparently smooth discs harbor unresolved gaps and rings, especially compact discs that are covered by a small number of resolution elements in current datasets. Even in the case of the DSHARP survey at 35 mas (≈ 5 AU; Andrews et al. 2018; Huang et al. 2018) there is evidence of only partially resolved substructure. Together these findings motivate the utility of a high resolution technique that can recover annular features on sub-beam scales and thus inform both theory and follow-up observations.

The ubiquity of annular substructures has prompted a

---

[*] E-mail: jmj51@ast.cam.ac.uk
[1] Notation: we use $\approx$ to mean 'approximately equal to' and $\sim$ to mean 'of order.'

variety of physical explanations for their origins, invoking forming and newly formed planets (e.g., Lin & Papaloizou 1979; Kley & Nelson 2012; Haffert et al. 2019; Isella et al. 2019) and/or disc processes such as preferential dust growth in localized regions (e.g., Pinilla et al. 2012; Zhang et al. 2015), opacity effects due to ice sublimation fronts (e.g., Zhang et al. 2016), internal photoevaporation (e.g., Clarke et al. 2001; Alexander et al. 2006), zonal flows (e.g., Flock et al. 2015), and the vertical shear instability (e.g., Flock et al. 2017). Highly accurate characterization of axisymmetric features can place constraints on these potential mechanisms for gap and ring creation, including the required dust properties (see applications in, e.g., Rosotti et al. 2016; Fedele et al. 2018; Clarke et al. 2018; Dullemond et al. 2018). Even in the presence of non-axisymmetric disc structure such as spirals, accurate recovery of the background radial profile is often an important first step, allowing isolation of the asymmetric features (Pérez et al. 2016; Meru et al. 2017). More accurate characterization of azimuthally averaged radial brightness profiles can thus aid in distinguishing the origins of both symmetric and asymmetric morphologies. While this could alternatively be achieved through higher resolution and/or deeper observations, extracting this information from existing datasets is more practically achievable.

This work will present a robust and accurate method to reconstruct an azimuthally averaged radial brightness profile by fitting the interferometric dataset. The challenge in this reconstruction (just as for the 2D brightness) is that the instrument response function resides in the Fourier domain, where the observations are an incomplete (sparse) representation of the source. Inverse Fourier transforming the visibilities to ultimately obtain a brightness profile thus requires some methodology to infer the unsampled visibilities. Three standard approaches for modeling protoplanetary disc observations are: use of the inverse modeling CLEAN nonlinear deconvolution algorithm (Högbom 1974; Clark 1980); forward modeling regularized maximum likelihood techniques that operate in the image plane, such as the maximum entropy method (e.g., Narayan & Nityananda 1986); and forward modeling by fitting a model directly to the visibilities.

Deprojecting and azimuthally averaging a CLEAN 2D model image is the most common technique used to obtain a disc's brightness profile. CLEAN constructs an empirical model for the sky brightness using a 'dirty image,' the inverse Fourier transform of the weighted visibilities, which is equivalent to the convolution of the sky brightness with the 'dirty beam' (the observational point spread function generated from the inverse Fourier transform of the weighted (u, v) distribution). Since the dirty beam has appreciable sidelobes, this introduces significant artifacts in the dirty image that CLEAN then attempts to remove. This is done by iteratively selecting the brightest pixel in the dirty image and adding a fraction of its amplitude to the model image in the form of a point source or Gaussian (Abrantes et al. 2009). This additional source is then convolved with the dirty beam and subtracted from the dirty image. The iteration proceeds until ideally the dirty image contains only noise. The resulting CLEAN model is then convolved with the 'CLEAN beam' (typically an elliptical Gaussian fit to the dirty beam's main lobe) to suppress the extrapolation of the model to scales below the beam (Cornwell et al. 1999) and give the final model image ('CLEAN image'). Impor-

tantly *this convolution spatially broadens and reduces in amplitude all features in the model image*. The effect is most severe for sub-beam structures but still alters even those resolved by the beam (see Sec. 3.1.1). Beam convolution thus places an intrinsic resolution limit on brightness profile extraction from the disc image. We will use several mock and real datasets to compare profile extraction from a CLEAN image to our model in this work.

An alternative approach to CLEAN is the forward modeling class of regularized maximum likelihood methods (RML). RML techniques for image reconstruction in radio interferometry have been used for decades (e.g., Ables 1974; Gull & Daniell 1978; Cornwell & Evans 1985) and have recently evolved rapidly, driven largely by their application to Event Horizon Telescope (EHT) observations and intent to apply them to Next Generation Very Large Array (ngVLA) observations (e.g., Lu et al. 2014; Honma et al. 2014; Chael et al. 2016; Johnson et al. 2017; Akiyama & Matthews 2019). The broad approach is to construct an image (among the infinite set of images consistent with the sparsely sampled visibilities) that simultaneously maximizes the likelihood (minimizes χ^2) and minimizes an additional constraint, such as the pixel brightness or image entropy (Narayan & Nityananda 1986; Event Horizon Telescope Collaboration et al. 2019). Importantly, RML methods have demonstrated success in attaining sub-beam resolution, including in the EHT observations of M87 (Event Horizon Telescope Collaboration et al. 2019).

A second alternative to CLEAN is to directly fit a forward model in the visibility domain. This avoids the resolution loss from beam convolution in CLEAN by inferring the unconvolved brightness distribution. Current forward models of this type (e.g., the `galario` code of Tazzari et al. 2018) require the user to specify a parametric functional form for the brightness profile (existing applications include the discs AS 209 and CI Tau; Fedele et al. 2017; Guzmán et al. 2018; Clarke et al. 2018). This approach advantageously allows one to motivate the profile's functional form using a physical model for the disc, such as the dust optical depth and temperature. However when the goal is simply to accurately fit the highest resolution information in a dataset, a parametric approach can face practical limitations.

These limitations are most evident with high resolution (baselines $\gtrsim 5$ Mλ) data, which almost all show fine structure in their visibilities at long baselines. The accuracy of the fit to this structure can strongly influence the recovered profile's identification and accurate characterization of sub-beam features. While there is no intrinsic resolution limit to a parametric fit of the visibilities, choosing a parametric model profile to closely match this long baseline structure is challenging. The functional form is often motivated by the CLEAN image, which for partially resolved, shallow or blended image features is not a trivial choice. Moreover the problem's dimensionality is typically high, making an exploration of multiple model profiles expensive. A disc with a single gap and ring has $15 - 20$ free parameters, and more structured discs can require >50. Parameter inference with a Markov Chain Monte Carlo (MCMC) sampler in this high dimensional space can take hundreds of CPU hours to converge. By comparison, a nonparametric approach can offer greater flexibility and speed to accurately fit long baseline, highly structured visibility distributions. We will use

the DSHARP observations of AS 209 (Andrews et al. 2018) to compare a parametric visibility fit to our nonparametric approach in this work.

To overcome the intrinsic resolution limit of CLEAN and the practical limitations of forward modeling in the visibility domain, we have developed a technique that fits the visibilities directly (avoiding beam convolution) and nonparametrically (affording the flexibility to fit complicated structure in the visibility distribution) to yield a reconstructed brightness profile. This empirical Bayes method, falling between a forward model with a fully explored posterior and a RML approach, imposes no assumptions on the functional form of the disc or its substructures, is autonomous (requiring no iterative, manual tuning of fit parameters), fast, and consistently achieves sub-beam fit resolution.

This work presents `frank`, an open source code to reconstruct the 1D radial brightness profile of an axisymmetric protoplanetary disc.[2] Sec. 2 details the code's methodology and assesses its prior sensitivities. Sec. 3 characterizes the model's performance, both intrinsic and as a function of data quality, using mock and real datasets. This includes application to low, moderate and high resolution datasets, and a detailed comparison to brightness profile extraction from a CLEAN image. Sec. 4 concludes by summarizing `frank`'s properties and outlining use cases for interferometric observations of protoplanetary discs.

2 MODEL

Our goal is to infer the true brightness $I_\nu(r)$ of a source under the assumption of azimuthal symmetry. To do this we will reconstruct $I_\nu(r)$ at a set of radial locations r_k by directly fitting the observed visibilities in the Fourier domain. This is possible by exploiting the properties of the *discrete Hankel transform (DHT)* detailed in Sec. 2.1 and 2.2.

We will show, however, that the direct fit is prone to find solutions with strong oscillations on small spatial scales. To solve this we introduce a Gaussian process in Sec. 2.3 to smooth the reconstructed profile. Specifically we will assume that the covariance matrix of the Gaussian process can be nonparametrically estimated from the data (visibilities) under the assumption that this matrix is diagonal in Fourier space. The free parameters (diagonal elements) of the matrix can be identified as the power spectrum of the reconstructed brightness profile. A similar approach, but based on log-normal priors, has previously been applied successfully to the more general problem of inferring 2D brightness distributions from radio observations (Sutter et al. 2013; Junklewitz et al. 2015, 2016; Greiner et al. 2016; Arras et al. 2018, 2019).

We begin by defining the visibility function of a source as sampled by an interferometer as the 2D Fourier transform of the source brightness (Cornwell et al. 1999),

$$V_\nu(u, v) = \int \int_S I_\nu(l, m) \exp(-2\pi i(ul + vm)) \, \mathrm{d}l\mathrm{d}m. \tag{1}$$

Here S indicates the region of the sky over which the integral is taken (which is assumed to be small, such that $|l^2 + m^2| \ll 1$); and $I_\nu(l, m)$ is the 2D sky brightness at real space antenna coordinates (l, m) and corresponding Fourier domain coordinates (u, v). (l, m) are measured in radians, (u, v) in the observing wavelength λ.

The necessary first step in fitting the visibilities is deprojection, correcting for the disc's on-sky inclination, position angle (rotation) and phase offset (departure from centering on the origin in the (u, v) plane). For mock observations in this work we will always consider face-on, phase-centered discs. For real observations we will use published geometry values (which we confirm by fitting a 2D Gaussian to the visibilities) to deproject the data prior to applying our model. Importantly, if a disc has an appreciable vertical thickness or if limb darkening from the optically thick surface is important, many of the assumptions underlying a deprojection approach such as fitting a 2D Gaussian to the visibilities would be invalidated.

Then for an azimuthally symmetric function (in our case a deprojected, assumed azimuthally symmetric disc) this 2D Fourier transform between the disc brightness and visibilities reduces to 1D as a Hankel transform with Bessel function kernels (Bracewell 2000; Thompson et al. 2017),

$$I_\nu(r) = \int V_\nu(q)J_0(2\pi qr)2\pi q \, \mathrm{d}q, \tag{2}$$

$$V_\nu(q) = \int I_\nu(r)J_0(2\pi qr)2\pi r \, \mathrm{d}r. \tag{3}$$

Here r is the radial coordinate in the disc, $q = \sqrt{u^2 + v^2}$ the baseline distance in the (u, v) plane, and J_0 the order 0 Bessel function of the first kind.

2.1 Representing the brightness profile as a Fourier-Bessel series

To evaluate Equations 2 – 3, we make use of their relation to Fourier-Bessel series via the DHT. For more information about the DHT see Baddour & Chouinard (2015)[3]; here we reproduce only the details necessary for our application in Sec. 2.2.

Imposing the assumption that the real space brightness profile $I_\nu(r) = 0$ beyond some radial distance R_{out}, or that the visibilities $V_\nu(q) = 0$ beyond some spatial frequency Q_{max}, enables the expansion of $I_\nu(r)$ or $V_\nu(q)$ in a Fourier-Bessel series. Respectively

$$I_\nu(r) = \sum_{k=1}^{\infty} \alpha_k J_0\left(\frac{j_{0k}r}{R_{\mathrm{out}}}\right), \tag{4}$$

$$V_\nu(q) = \sum_{k=1}^{\infty} \beta_k J_0\left(\frac{j_{0k}q}{Q_{\mathrm{max}}}\right), \tag{5}$$

where j_{0k} is a scalar representing the kth root of $J_0(r)$, i.e., $J_0(j_{0k}) = 0$.

The coefficients α_k in Equation 4 can be computed via

---

[3] N.B. The definition of the Hankel transform used here differs from that in Baddour & Chouinard (2015) by factors of 2π.

the orthogonality relationship of Bessel functions with subscripts k and j,

$$\int_0^{R_{\text{out}}} J_0\left(\frac{j_{0k}r}{R_{\text{out}}}\right) J_0\left(\frac{j_{0j}r}{R_{\text{out}}}\right) 2\pi r \mathrm{d}r = \pi R_{\text{out}}^2 J_1^2(j_{0j})\delta_{jk}, \tag{6}$$

where δ_{ij} is the Kronecker δ and J_1 the first order Bessel function. Substituting Equation 4 into Equation 6,

$$\begin{aligned}
\int_0^{R_{\text{out}}} & I_\nu(r) J_0\left(\frac{j_{0k}r}{R_{\text{out}}}\right) 2\pi r \mathrm{d}r \\
&= \sum_{j=1}^{\infty} \alpha_j \int_0^{R_{\text{out}}} J_0\left(\frac{j_{0k}r}{R_{\text{out}}}\right) J_0\left(\frac{j_{0j}r}{R_{\text{out}}}\right) 2\pi r \mathrm{d}r \\
&= \alpha_k \pi R_{\text{out}}^2 J_1^2(j_{0k}).
\end{aligned} \tag{7}$$

Noting that the left-hand side of this is just the Hankel transform of I_ν, the brightness profile can be written entirely in terms of its visibilities at a specific set of spatial frequencies $q_k = j_{0k}/(2\pi R_{\text{max}})$ and the Fourier-Bessel series coefficients α_k. Analogously the visibility at any q can be computed in terms of the brightness at the set of radial locations (collocation points) $r_k = j_{0k}/(2\pi Q_{\text{max}})$ and the Fourier-Bessel series coefficients β_k,

$$\alpha_k = \frac{1}{\pi R_{\text{out}}^2 J_1^2(j_{0k})} V_\nu\left(\frac{j_{0k}}{2\pi R_{\text{out}}}\right), \tag{8}$$

$$\beta_k = \frac{1}{\pi Q_{\text{max}}^2 J_1^2(j_{0k})} I_\nu\left(\frac{j_{0k}}{2\pi Q_{\text{max}}}\right). \tag{9}$$

In practice we must truncate the infinite sums in Equation 4 and Equation 5 to a finite value N. From Equation 4, the brightness profile then becomes entirely determined by spatial frequencies below some $2\pi q = j_{0N_+}/R_{\text{out}}$, where j_{0N_+} is j_{0k} for $k = N + 1$. Similarly the visibilities are entirely determined by radii smaller than $2\pi r = j_{0N_+}/Q_{\text{max}}$. Choosing $2\pi Q_{\text{max}} = j_{0N_+}/R_{\text{out}}$ then produces the DHT (Baddour & Chouinard 2015). The rules for the backward (visibility space → real space) and forward (real space → visibility space) transforms of the DHT are

$$I_k = \frac{j_{0N_+}}{2\pi R_{\text{out}}^2} \sum_{j=1}^N Y_{kj} V_j, \tag{10}$$

$$V_k = \frac{2\pi R_{\text{out}}^2}{j_{0N_+}} \sum_{j=1}^N Y_{kj} I_j, \tag{11}$$

where

$$Y_{kj} = \frac{2}{j_{0N_+} J_1^2(j_{0j})} J_0\left(\frac{j_{0k}j_{0j}}{j_{0N_+}}\right). \tag{12}$$

The intensities $I_k = I_\nu(r_k)$ and visibilities $V_k = V_\nu(q_k)$ are evaluated at the collocation points of the Fourier-Bessel series in real and visibility space respectively,

$$r_k = R_{\text{out}} j_{0k}/j_{0N_+}, \tag{13}$$

$$q_k = j_{0k}/(2\pi R_{\text{out}}). \tag{14}$$

We illustrate the correspondence between the brightness profile and visibilities using the example of a Gaussian brightness profile in **Fig. 1**. A small number of collocation points ($\lesssim 10$, corresponding to the same number of terms in the Bessel series) yields, via the DHT, a visibility profile $V_\nu(q)$ that is in good agreement with the analytic Hankel

transform of the input profile up to frequencies $q \sim Q_{\text{max}}$. In practice to account for more complicated profiles we use $100 - 300$ collocation points.

It is convenient to absorb the normalization coefficients from Equations 10 – 11 into backward and forward transform matrices respectively[4],

$$\mathbf{Y}_{\text{b}} = \frac{j_{0N_+}}{2\pi R_{\text{out}}^2} \mathbf{Y}, \tag{15}$$

$$\mathbf{Y}_{\text{f}} = \frac{2\pi R_{\text{out}}^2}{j_{0N_+}} \mathbf{Y}, \tag{16}$$

which obey

$$\mathbf{Y}_{\text{b}}\mathbf{Y}_{\text{f}} = \mathbf{Y}\mathbf{Y} \approx \mathbf{I}, \tag{17}$$

where $\mathbf{I}$ is the identity matrix. The last approximation is exact only for $N \to \infty$, though the error is small at modest N; for $N > 30$ the largest error is $<10^{-7}$. In the code the impact is even less significant because only the forward transform matrices are used explicitly.

These matrices can be used to specify the transformation rules for vectors,

$$\boldsymbol{f} = \mathbf{Y}_{\text{b}}\tilde{\boldsymbol{f}}, \tag{18}$$

$$\tilde{\boldsymbol{f}} = \mathbf{Y}_{\text{f}}\boldsymbol{f}, \tag{19}$$

where we explicitly use a tilde (e.g., $\tilde{\boldsymbol{f}}$) to distinguish Fourier domain quantities from a real space vector (e.g., $\boldsymbol{f}$). For the visibilities $\boldsymbol{V}$ we will drop the tilde.

It will also be useful to define the transformation rules for a covariance matrix $\mathbf{S}$ and its inverse. These can be derived from the equivalence of scalars in real and visibility space, $\boldsymbol{f}^T \mathbf{S}^{-1} \boldsymbol{f} = \tilde{\boldsymbol{f}}^T \tilde{\mathbf{S}}^{-1} \tilde{\boldsymbol{f}}$, where $\tilde{\mathbf{S}}$ is the visibility space representation of the covariance matrix. From this relation we have

$$\tilde{\mathbf{S}}^{-1} = \mathbf{Y}_{\text{b}}^T \mathbf{S}^{-1} \mathbf{Y}_{\text{b}}, \tag{20}$$

$$\tilde{\mathbf{S}} = \mathbf{Y}_{\text{f}}^T \mathbf{S} \mathbf{Y}_{\text{f}}. \tag{21}$$

2.2 Fitting the visibilities using the discrete Hankel transform

Using the Fourier-Bessel series and DHT from Sec. 2.1, we now develop a method to reconstruct a disc's brightness profile given a set of N_{vis} visibilities. To keep the notation succinct we will denote the visibilities by the vector $\boldsymbol{V}$, with associated baselines $\boldsymbol{q}$ and corresponding weights $\boldsymbol{w}$. For a radio interferometer it is reasonable to assume that the noise on each visibility is drawn from an independent Gaussian distribution with variance $\sigma_j^2 = 1/w_j$, such that the covariance matrix of the noise is $\mathbf{N} = \text{diag}(1/\boldsymbol{w})$. That is, $\mathbf{N}$ has $1/w_j$ along the diagonal and is 0 otherwise.

Fixing R_{out} (to a distance beyond the edge of the disc) and choosing the number of brightness points N fixes the radial collocation points r_k. We can then use the Fourier-Bessel series representation (Equation 5) to link the observed visibilities $\boldsymbol{V}$ to $I_\nu(r_k)$, which we seek to infer. The likelihood for $\boldsymbol{V}$ is a Gaussian,

$$\mathcal{L}_{\text{G}} = P(\boldsymbol{V}|\boldsymbol{I}_\nu) = \mathcal{G}(\boldsymbol{V} - \mathbf{H}(\boldsymbol{q})\boldsymbol{I}_\nu, \mathbf{N}), \tag{22}$$

---

[4] Notation: We use boldface for matrix quantities, e.g., $\mathbf{Y}=Y_{kj}$.
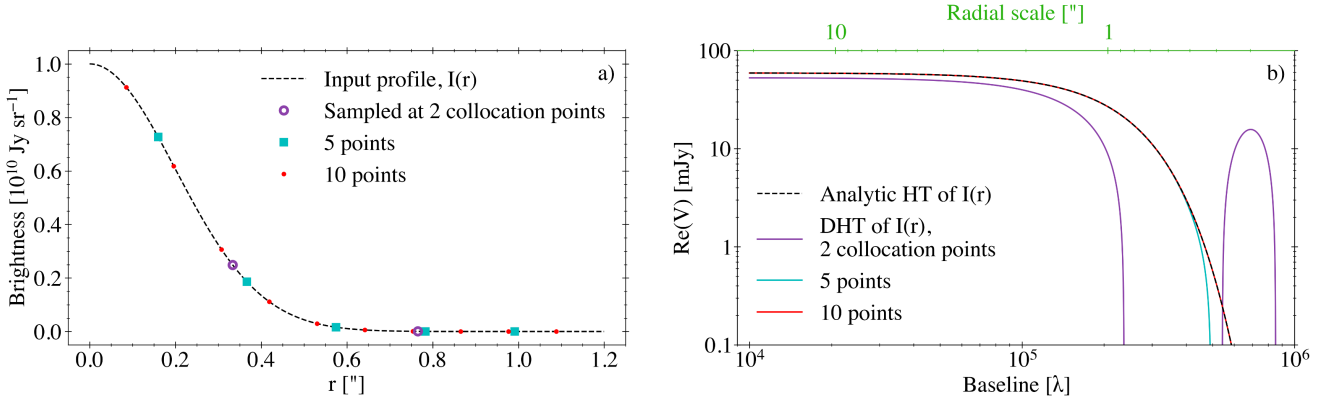
Figure 1. Fourier-Bessel series representation of a brightness profile
a) An input mock brightness profile for a Gaussian disc, with samplings at an increasing number of radial collocation points and $R_{out} = 1.2''$.
b) The discrete Hankel transform (DHT) of a Fourier-Bessel series representation of the input profile at these sets of collocation points, showing that an approximation with a small number of points can closely match the analytic Hankel transform (HT) of the input profile. The second x-axis shows the spatial scale corresponding to a given baseline, $r_{scale} = 1/q$.

where generically $\mathcal{G}(M, \Sigma)$ refers to a multidimensional Gaussian with mean M and covariance Σ. The vector I_ν is the brightness at the radial collocation points (i.e., it has the components $I_\nu(r_k)$), and we have introduced the $N_{vis} \times N$ matrix $H(q)$, defined by the components

$$H_k(q_j) = \frac{4\pi R_{out}^2}{j_{0N_+}^2 J_1^2(j_{0k})} J_0\left(2\pi q_j R_{out} \frac{j_{0k}}{j_{0N_+}}\right), \qquad (23)$$

which comes from the Fourier-Bessel series expansion.

One way to derive $I_\nu(r_k)$ would be to maximize $\mathcal{L}_G$. The solution would be

$$I_\nu = M^{-1} j, \qquad (24)$$

where

$$M = H(q)^T N^{-1} H(q), \qquad (25)$$

$$j = H(q)^T N^{-1} V. \qquad (26)$$

The model's dependence on the visibility data enters entirely through M and j. Note that the construction of M scales as $O(N^2 N_{vis})$, and the construction of j scales as $O(N N_{vis})$, while the solution of these equations and subsequent expressions in Sec. 2.3 scale as $O(N^3)$, where we recall N is the number of collocation points. Because M and j are constructed using all unbinned visibilities, the code does not regrid the visibilities onto the spatial frequency collocation points; we are evaluating the full set of observed visibilities. A value of $N \approx 100 - 300 \ll N_{vis}$ is sufficient to fit data at the highest resolutions of current interferometers.

The problem with maximizing $\mathcal{L}_G$ directly is that for sufficiently large N, M will become singular. This occurs because the relationship between nearby points in the brightness profile is determined by high frequency components. For sufficiently large N these components are either not present in, or are poorly constrained by, the visibilities. Yet the requirement that N is small enough that we are able to invert M will be too restrictive, limiting our ability to accurately fit high signal-to-noise (SNR), short baseline data. We would need to fit a large number of data points under the constraint that the profile is smooth on sufficiently small scales.

Fitting a brightness profile with a reasonable number of radial points therefore requires regularizing the solution on small spatial scales.

Fig. 2 shows how attempting to fit a brightness profile without regularization yields numerical instability in M at a modest number of collocation points, $N = 30$. While at $N = 25$ the fit is stable, the resulting brightness profile is undersampled, with variations between adjacent radial collocation points. Adding more points to the unregularized fit causes these oscillations to increase in amplitude and frequency, with the visibility domain fit in Fig. 2(b) having erroneously high amplitude near and beyond the data's longest baselines. In Sec. 2.3 we will describe how the fit can be regularized using a nonparametric Gaussian process model. The regularized fit with `frank` shown in Fig. 2(a) is smooth and insensitive to the number of collocation points (we show the case for $N = 200$), yielding an accurate recovery of the input profile. The visibility domain fit in Fig. 2(b) correspondingly decreases in amplitude at the longest baselines and beyond (note the data are noise-dominated beyond ≈ 1.2 Mλ).

This difference in behavior between the unregularized and `frank` fits is a consequence of the correlations between radial collocation points in each model. In the unregularized fit, adjacent points are almost perfectly anticorrelated for $N = 30$ in Fig. 2(c), inducing strong brightness profile oscillations. By contrast the regularized `frank` fit introduces a positive correlation between adjacent points in Fig. 2(d), damping oscillations in the recovered brightness profile.

2.3 Regularizing the fit using a nonparametric Gaussian process

Regularization corresponds to an a priori assumption that the brightness should be highly correlated at adjacent points and more weakly correlated at distant points. This assumption is well suited to the framework of a Gaussian process, in which the prior on I_ν is a Gaussian,

$$P(I_\nu | p) = \mathcal{G}(I_\nu, S(p)), \qquad (27)$$
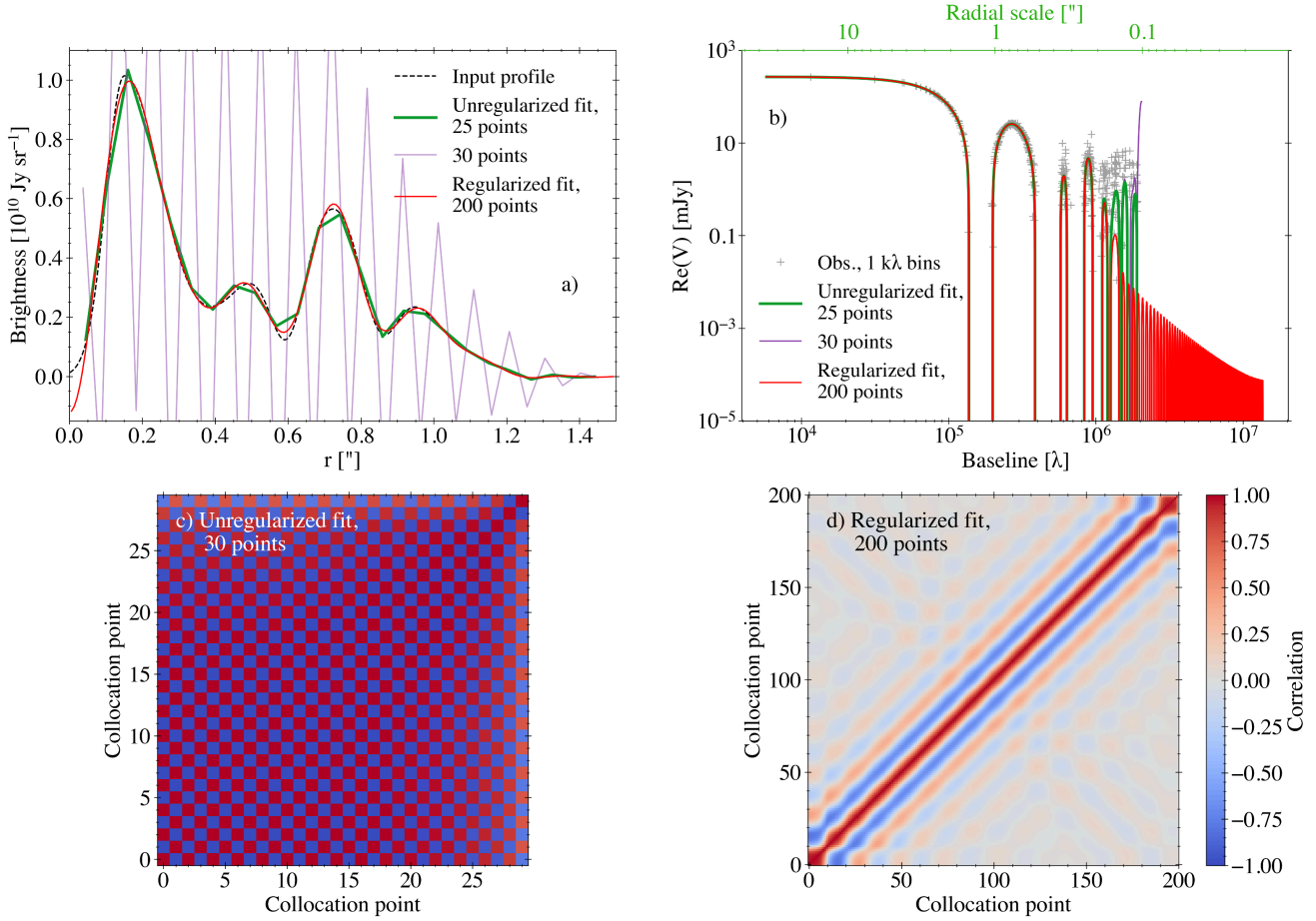
Figure 2. Fit regularization

a) For a multi-Gaussian mock disc observed with the ALMA C43-6 configuration (beam FWHM $0.13 \times 0.17''$, Briggs=0.5; see Table 2), a fit with regularization and using 200 collocation points. Fits derived from Equation 24 without regularization are shown for comparison, demonstrating instability at only 30 collocation points.

b) Visibilities for the mock observation in (a) and fits corresponding to the brightness profiles in (a). The unregularized fits place erroneously high (noise) power beyond the data's longest baseline, while the regularized fit yields a more reasonable prediction for power on these unobserved scales.

c) 30×30 correlation matrix for the 30-point unregularized fit in (a), showing that the high amplitude oscillations in (b) are a result of almost perfectly strong anticorrelation between adjacent points.

d) 200×200 posterior correlation matrix for the regularized fit in (a), with the regularization providing stability by damping correlations. This in turn prevents regions of erroneously high power in (b) and thus spurious oscillations in (a).

where $\mathbf{S}(\boldsymbol{p})$ is the prior covariance in the real space brightness profile at the radial collocation points. We explicitly specify that the covariance structure $\mathbf{S}(\boldsymbol{p})$ has some dependence on a set of parameters $\boldsymbol{p}$, which we will relate to an estimate of the power spectrum based on the DHT of $\boldsymbol{I}_\nu$, $\tilde{\boldsymbol{I}}_\nu^2 = (\mathbf{Y}_{\mathrm{f}} \boldsymbol{I}_\nu)^2$. Given $\boldsymbol{p}$, the posterior probability for $\boldsymbol{I}_\nu$ can be used to reconstruct the brightness,

$$P(\boldsymbol{I}_\nu | \boldsymbol{V}, \boldsymbol{p}) = \frac{P(\boldsymbol{V} | \boldsymbol{I}_\nu, \boldsymbol{p}) P(\boldsymbol{I}_\nu | \boldsymbol{p})}{P(\boldsymbol{V} | \boldsymbol{p})} \tag{28}$$

$$= \frac{\mathcal{G}(\boldsymbol{V} - \mathbf{H}(\boldsymbol{q}) \boldsymbol{I}_\nu, \mathbf{N}) \, \mathcal{G}(\boldsymbol{I}_\nu, \mathbf{S}(\boldsymbol{p}))}{P(\boldsymbol{V} | \boldsymbol{p})}. \tag{29}$$

The numerator here is the product of two Gaussians, which is also a Gaussian, and has covariance $\mathbf{D}$ and mean $\boldsymbol{\mu}$,

$$\mathbf{D} = \left(\mathbf{M} + \mathbf{S}(\boldsymbol{p})^{-1} \right)^{-1},$$

$$\boldsymbol{\mu} = \mathbf{D} \, \boldsymbol{j}. \tag{30}$$

Explicitly,

$$P(\boldsymbol{I}_\nu | \boldsymbol{V}, \boldsymbol{p}) \propto \mathcal{G}(\boldsymbol{I}_\nu - \boldsymbol{\mu}, \mathbf{D}), \tag{31}$$

which we will use to infer $\boldsymbol{I}_\nu$ given $\boldsymbol{p}$. The remaining challenge is how to specify $\mathbf{S}(\boldsymbol{p})$. Typically in a Gaussian process the covariance structure is parameterized in terms of a simple function, such as a Gaussian with some length scale (Rasmussen & Williams 2006). This length scale could then be optimized or better yet marginalized over. We follow an alternative approach, the empirical Bayes method, in which

we use a nonparametric form for the covariance matrix that can be estimated from the visibilities simultaneously with the brightness profile. This approach follows the work of Oppermann et al. (2013); see also Enßlin & Frommert (2011).

We make the ansatz that the prior covariance matrix is diagonal in visibility space, i.e.,

$$\tilde{\mathbf{S}}(\boldsymbol{p}) = \mathbf{Y}_f^T \mathbf{S}(\boldsymbol{p}) \mathbf{Y}_f = \text{diag}(\boldsymbol{p}), \tag{32}$$

thus

$$\mathbf{S}(\boldsymbol{p}) = \mathbf{Y}_b^T \text{diag}(\boldsymbol{p}) \mathbf{Y}_b, \tag{33}$$

where we have now defined the parameters $\boldsymbol{p}$ as the diagonal elements of the visibility space representation of the covariance matrix (with the off-diagonal elements set to zero). In the code, we construct $\mathbf{S}(\boldsymbol{p})^{-1}$ directly from $1/\boldsymbol{p}$ and $\mathbf{Y}_f$.

To understand the effect of this prior, we consider $\tilde{\boldsymbol{I}}_\nu$ – the Fourier space representation of $\boldsymbol{I}_\nu$ – which is equivalent to the predicted visibility at the spatial frequency collocation points q_k. In visibility space the prior takes the form

$$\log P(\boldsymbol{I}_\nu | \mathbf{S}(\boldsymbol{p})) \equiv \log P(\tilde{\boldsymbol{I}}_\nu | \tilde{\mathbf{S}}(\boldsymbol{p}))$$
$$= -\frac{1}{2} \boldsymbol{I}_\nu^T \mathbf{S}(\boldsymbol{p})^{-1} \boldsymbol{I}_\nu - \frac{1}{2} \log |2\pi \mathbf{S}(\boldsymbol{p})|$$
$$= -\frac{1}{2} \sum_k \left(\frac{\tilde{I}_{\nu,k}^2}{p_k} + \log p_k \right) + \text{const}, \tag{34}$$

where $\tilde{I}_{\nu,k}$ and p_k refer to the kth element of $\tilde{\boldsymbol{I}}_\nu$ and $\boldsymbol{p}$. The last line follows from the definition of $\tilde{\boldsymbol{I}}_\nu$ and the relation between the determinant of a matrix product and the product of determinants,

$$|2\pi \mathbf{S}(\boldsymbol{p})| = |\mathbf{Y}_b^T| \cdot |2\pi \text{diag}(\boldsymbol{p})| \cdot |\mathbf{Y}_b| = |\mathbf{Y}_b|^2 \prod_k 2\pi p_k. \tag{35}$$

From Equation 34 we see that if $\tilde{I}_{\nu,k}^2$ (the power in the brightness profile on a scale k) is large relative to p_k, the prior probability will be small. Thus the prior acts to suppress power on scales where p_k is small. We examine the prior's effect on the reconstructed brightness profile in greater depth in Sec. 2.5.

2.4 Jointly inferring the brightness profile and power spectrum parameters

Because we do not know a priori the optimal choice for $\boldsymbol{p}$, reconstructing the brightness profile is now a problem of jointly inferring $\boldsymbol{I}_\nu$ and $\boldsymbol{p}$. The joint posterior probability $P(\boldsymbol{I}_\nu, \boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta})$ is constructed using the posterior for $\boldsymbol{I}_\nu$ given $\boldsymbol{p}$, $P(\boldsymbol{I}_\nu | \boldsymbol{V}, \boldsymbol{p})$, and a prior probability distribution for $\boldsymbol{p}$, $P(\boldsymbol{p}, \boldsymbol{\beta})$, via

$$P(\boldsymbol{I}_\nu, \boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta}) = P(\boldsymbol{I}_\nu | \boldsymbol{V}, \boldsymbol{p}) P(\boldsymbol{p} | \boldsymbol{\beta}). \tag{36}$$

Here we have noted explicitly the dependence of $P(\boldsymbol{p} | \boldsymbol{\beta})$ on a set of hyperparameters $\boldsymbol{\beta} = \{\alpha, p_0, w_{\text{smooth}}\}$, which also introduces the dependence on $\boldsymbol{\beta}$ into the posterior probability $P(\boldsymbol{I}_\nu, \boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta})$. The set of parameters $\boldsymbol{\beta}$ will be held fixed in any given inference of $\boldsymbol{I}_\nu$ and $\boldsymbol{p}$. We will refer to $P(\boldsymbol{p} | \boldsymbol{\beta})$ as the hyperprior to distinguish it from $P(\boldsymbol{I}_\nu | \boldsymbol{p})$. We define the components of $\boldsymbol{\beta}$ below.

To estimate the parameters of the covariance matrix $\boldsymbol{p}$ using the data, our general approach will be to produce small values for $\boldsymbol{p}$ on scales that are unconstrained by the

data, in order to suppress them, but otherwise allow $\boldsymbol{p}$ to be sufficiently large that the reconstructed brightness profile is controlled by the data. To achieve this we specify $P(\boldsymbol{p} | \boldsymbol{\beta})$ as the product of a spectral smoothness term and inverse Gamma functions,

$$P(\boldsymbol{p} | \boldsymbol{\beta}) = P_{\text{smooth}}(\boldsymbol{p} | w_{\text{smooth}}) \prod_{k=1}^N \frac{1}{p_0 \Gamma(\alpha-1)} \left(\frac{p_k}{p_0} \right)^{-\alpha} \exp\left(-\frac{p_0}{p_k} \right). \tag{37}$$

The exponential part of the inverse Γ function disfavors values of $p_k < p_0$, while the power law disfavors $p_k \gg p_0$ for $\alpha > 1$. Neglecting the spectral smoothness hyperprior, the limit $\alpha \to 1$ and $p_0 \to 0$ yields a Jeffreys prior (flat in log space). We typically choose a small but nonzero value of p_0 (e.g., 10^{-15} Jy^2) for practicality; low p_0 allows frank to find a solution with very low power on scales unconstrained by the data, strongly regularizing those scales. Though we do not want p_0 to be arbitrarily small, as this leads to numerical instability when evaluating Equation 30.

The spectral smoothness hyperprior follows Oppermann et al. (2013) and is included for two reasons. It first prevents regions of artificially low power arising from narrow gaps in the visibilities and at unconstrained scales beyond the data's longest baseline, ensuring the brightness profile does not exhibit artificially high correlation at the corresponding spatial scales. Secondly it introduces a coupling between adjacent points in the power spectrum. This has the effect of 'averaging' the squared visibility amplitude over a range of scales, suppressing the impact of noise on the power spectrum. Overall we have not found the brightness reconstruction to be highly sensitive to the inclusion of the smoothing hyperprior, which takes the form

$$P_{\text{smooth}}(\boldsymbol{p} | w_{\text{smooth}} \equiv 1/\sigma_s^2) \propto$$
$$\exp\left(-\frac{1}{2\sigma_s^2} \int d\log(q) \left(\frac{\partial^2 \log(p)}{\partial \log(q)^2} \right)^2 \right). \tag{38}$$

This hyperprior penalizes power spectra with large second derivatives in log space, i.e., those that deviate from a power law (a straight line in log space). σ_s, which is parameterized in terms of w_{smooth}, controls the allowed amount of variation (departure from a power law) in the power spectrum at a given q. We implement this hyperprior using a numerical estimate of $\partial^2 \log(p)/\partial \log(q)^2$, which can be written in the form

$$P_{\text{smooth}}(\boldsymbol{p} | w_{\text{smooth}}) \propto \exp\left(-\frac{1}{2} \log(\boldsymbol{p})^T \frac{\mathbf{T}}{\sigma_s^2} \log(\boldsymbol{p}) \right), \tag{39}$$

where $\mathbf{T}$ is a constant, pentadiagonal matrix that depends only on the spatial frequency collocation points q_k. For the exact form of $\mathbf{T}$, see Appendix A.

With $P_{\text{smooth}}(\boldsymbol{p} | w_{\text{smooth}})$ specified, we now have a form for $P(\boldsymbol{p} | \boldsymbol{\beta})$. Ideally we would proceed to obtain the posterior for $\boldsymbol{I}_\nu$ by marginalizing over $\boldsymbol{p}$, but the high dimensionality of the parameter space makes this impractical. We instead maximize $P(\boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta})$ to obtain and use the maximum a posteriori value $\boldsymbol{p}_{\text{MAP}}$ as the prior on $\boldsymbol{I}_\nu$. The marginal posterior $P(\boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta})$ is obtained from $P(\boldsymbol{I}_\nu, \boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta})$ by integrating over all $\boldsymbol{I}_\nu$, i.e., $\int P(\boldsymbol{I}_\nu, \boldsymbol{p} | \boldsymbol{V}, \boldsymbol{\beta}) \, d\boldsymbol{I}_\nu$. Since $P(\boldsymbol{I}_\nu | \boldsymbol{V}, \boldsymbol{p})$ is a multivariate Gaussian, this can be done analytically (e.g., Appendix A of Rasmussen & Williams 2006) as in Oppermann

et al. (2013),

$$\log P(\boldsymbol{p}\,|\,\boldsymbol{V},\boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{j}^T\mathbf{D}\,\boldsymbol{j} + \frac{1}{2}\log|\mathbf{D}| - \frac{1}{2}\log|\mathbf{S}(\boldsymbol{p})|$$
$$- \sum_k \left[(\alpha-1)\log p_k + \frac{p_0}{p_k}\right] - \frac{1}{2}\log(\boldsymbol{p})^T\frac{\mathbf{T}}{\sigma_s^2}\log(\boldsymbol{p})$$
$$+ \text{const.} \tag{40}$$

(see also Enßlin & Frommert 2011). Finding the maximum entails finding the location where the derivative of Equation 40 with respect to $\log p_k$,

$$\frac{\mathrm{d}\log P(\boldsymbol{p}\,|\,\boldsymbol{V},\boldsymbol{\beta})}{\mathrm{d}\log p_k} = \frac{1}{2p_k}\left[\left(\mathbf{Y}_{\mathrm{f}}(\boldsymbol{\mu}\boldsymbol{\mu}^T + \mathbf{D})\mathbf{Y}_{\mathrm{f}}^T\right)_{kk}\right] - \frac{1}{2}$$
$$- \left[(\alpha-1) - \frac{p_0}{p_k}\right] - \left(\frac{\mathbf{T}}{\sigma_s^2}\log\boldsymbol{p}\right)_k, \tag{41}$$

is zero. We find the maximum using the fixed point iteration

$$(\mathbf{I}+\frac{\mathbf{T}}{\sigma_s^2})\log\boldsymbol{p}^{\mathrm{new}} = \log\boldsymbol{p}$$
$$+ \frac{1}{p_i}\left[p_0 + \frac{1}{2}\mathrm{diag}\left(\mathbf{Y}_{\mathrm{f}}(\boldsymbol{\mu}\boldsymbol{\mu}^T + \mathbf{D})\mathbf{Y}_{\mathrm{f}}^T\right)\right] - \left[(\alpha-1)+\frac{1}{2}\right] \tag{42}$$

(recall that $\mathbf{I}$ is the identity matrix). Here the application of the $\mathrm{diag}(\mathbf{M})$ operator to a matrix should be understood as selecting the vector formed from the diagonal elements of that matrix. At each iteration $\boldsymbol{\mu}$ and $\mathbf{D}$ are computed using $\boldsymbol{p}$ from the previous iteration, and the linear system is solved using a sparse linear solver. Each iteration requires $O(N^3)$ operations. The iterations are terminated when the relative change in $\boldsymbol{p}$ is $< 10^{-3}$. We have confirmed with tests that the final solution is not sensitive to the initial choice of $\boldsymbol{p}$; we use as an initialization a power law with slope of -2 to coarsely match the typical decline in visibility amplitude as a function of baseline in high resolution ALMA observations.

For the final reconstructed brightness profile, we use our best-fitting (maximum a posteriori) values for the power spectrum coefficients $\boldsymbol{p}_{\mathrm{MAP}}$ in Equation 30. This provides our estimate for the profile's mean $\boldsymbol{\mu}$ and covariance $\mathbf{D}$. In general $\mathrm{diag}(\mathbf{D})$ will underestimate the uncertainty on the brightness at each collocation point, as discussed in Sec 2.7. We summarize the overall model framework diagrammatically in **Fig. 3**.

2.5 $\boldsymbol{p}_{\mathrm{MAP}}$ and its use as a prior on I_ν

The maximum a posteriori power spectrum $\boldsymbol{p}_{\mathrm{MAP}}$ tends to one of two limiting forms, depending primarily on the visibility SNR at each spatial frequency. To understand this we consider the converged maximum a posteriori values (i.e., $\log\boldsymbol{p}^{\mathrm{new}} = \log\boldsymbol{p}$), neglecting the spectral smoothness hyperprior for now ($w_{\mathrm{smooth}} \to 0$).

First, at spatial frequencies where the visibility SNR is sufficiently high, the term involving the mean $\boldsymbol{\mu}$ in Equation 42 dominates and thus

$$\boldsymbol{p}_{\mathrm{MAP}} = \frac{(\mathbf{Y}_{\mathrm{f}}\boldsymbol{\mu})^2}{1 + 2(\alpha-1)}. \tag{43}$$

In this case the power spectrum is set by an estimate using the DHT of the mean brightness profile, which is approximately the square of the visibility amplitude (its power). This demonstrates the aforementioned association between
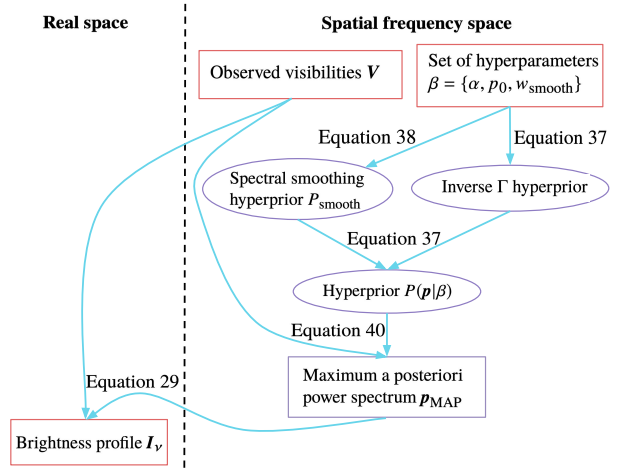
Figure 3. Model framework
Diagram of the probabilistic model framework in `frank`. Squares represent variables, purple indicates the quantity is either a hyperprior or a prior (note $\boldsymbol{p}_{\mathrm{MAP}}$ is an inferred variable that is used as a prior). The set of hyperparameters $\boldsymbol{\beta}$ determine the hyperpriors placed on the power spectrum reconstruction from $\boldsymbol{V}$, yielding $\boldsymbol{p}_{\mathrm{MAP}}$. This is then used as a prior for the reconstruction of I_ν from $\boldsymbol{V}$.

$\boldsymbol{p}_{\mathrm{MAP}}$ and the power spectrum and provides the justification for calling $\boldsymbol{p}_{\mathrm{MAP}}$ the power spectrum.

Conversely, where the visibility SNR is sufficiently low, $\boldsymbol{p}_{\mathrm{MAP}}$ is small enough that the $\mathbf{S}(\boldsymbol{p})$ term in $\mathbf{D}$ dominates. In this case

$$\boldsymbol{p}_{\mathrm{MAP}} = \frac{p_0 + \frac{1}{2}\mathrm{diag}\left(\mathbf{Y}_{\mathrm{f}}\mathbf{S}(\boldsymbol{p})\mathbf{Y}_{\mathrm{f}}\right)}{\alpha - 1 + 1/2} = \frac{p_0 + \boldsymbol{p}_{\mathrm{MAP}}/2}{\alpha - 1 + 1/2} = \frac{p_0}{\alpha - 1}. \tag{44}$$

The SNR threshold that separates these behaviors depends on α (Appendix B); at $\alpha = 1$ the threshold is at SNR $= 1$, and the threshold increases as α increases (because for larger α, the inverse Γ hyperprior decays faster with increasing p_k). Thus given the same data, a larger α will cause p_k to be pulled more strongly downward toward p_0.

Secondary to the effect of the visibility SNR, the spectral smoothness hyperprior modifies the power spectrum. Fig. 4(a) and (d) (presented below) demonstrate that increasing w_{smooth} reduces structure in the power spectrum, driving its shape toward a power law. A similar effect is evident in Fig. 5 (presented in Sec. 2.6), where the oscillations in the power spectrum under a Jeffrey's prior (using $w_{\mathrm{smooth}} = 0$) are large compared with those generated under $w_{\mathrm{smooth}} = 10^{-4}$. This demonstrates that in regions of low SNR at long baseline, the constraint of a smooth power spectrum can dominate the inverse Γ hyperprior's preference for low p_i. Though this is typically isolated to the case of $\alpha = 1$, because the inverse Γ hyperprior does not damp the power spectrum coefficients under this choice.

Given a form for the maximum a posteriori power spectrum $\boldsymbol{p}_{\mathrm{MAP}}$, we now consider its effects as a prior on the reconstructed brightness profile I_ν. **Fig. 4** shows multiple power spectra generated using `frank` fits to the mock brightness profile in Fig. 2. Under different ALMA configurations and values of w_{smooth}, the power spectra in Fig. 4(a) and (d)
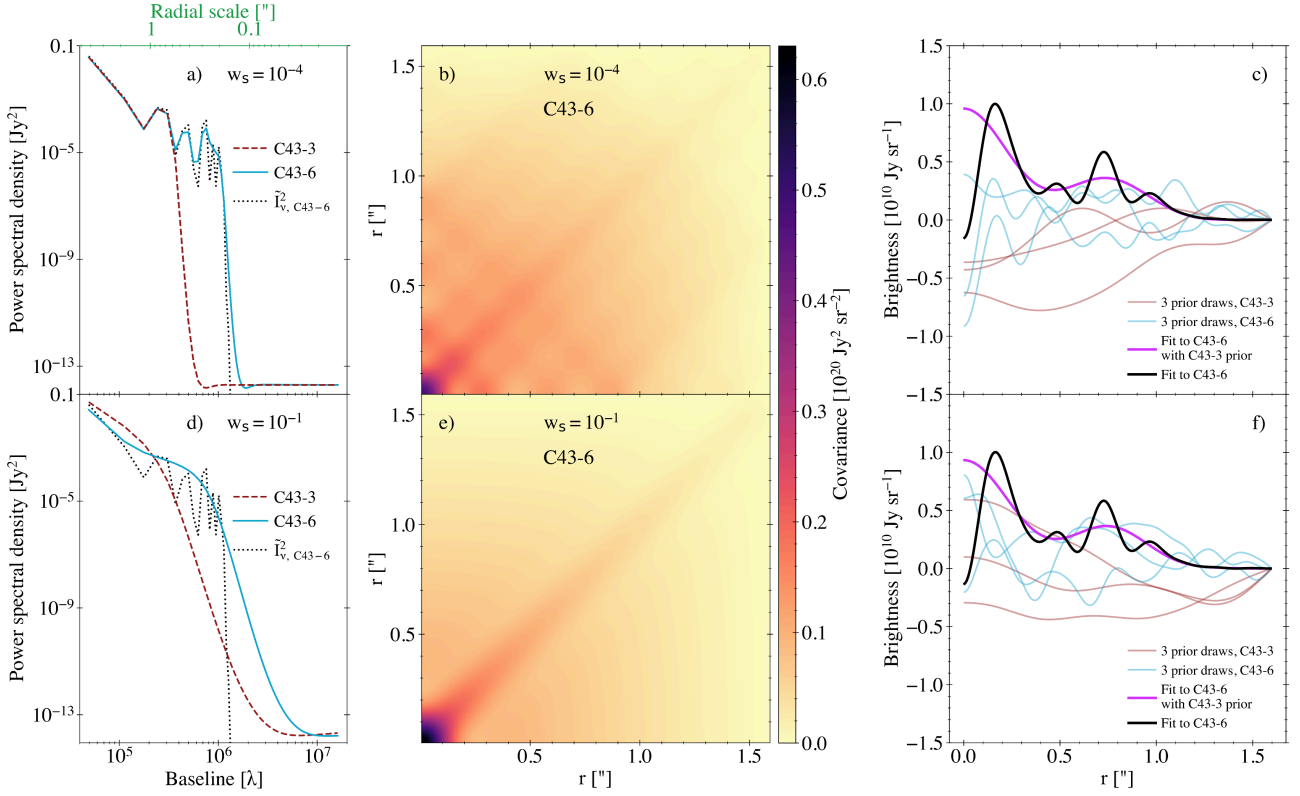
Figure 4. Effect of the hyperparameter w_{smooth}

a) Maximum a posteriori power spectra $\boldsymbol{p}_{MAP}$ for a multi-Gaussian mock disc observed separately with the ALMA C43-3 (synthesized beam FWHM $0.59 \times 0.70''$, Briggs=0.5) and C43-6 (beam FWHM $0.13 \times 0.17''$, Briggs=0.5) configurations (see Table 2). These $\boldsymbol{p}_{MAP}$ are obtained for each dataset using $P(\boldsymbol{p}|\boldsymbol{V},\boldsymbol{\beta})$, the frank posterior marginalized over all realizations of $\boldsymbol{I}_\nu$, under the hyperparameters $\boldsymbol{\beta} = \{\alpha, p_0, w_{smooth}\} = \{1.05, 10^{-15} \text{ Jy}^2, 10^{-4}\}$. For comparison the power spectrum estimate $\tilde{I}^2_{\nu, \text{ C43-6}} = (\mathbf{Y}_f \boldsymbol{I}_{\nu, \text{ C43-6}})^2$ based on the frank fitted brightness profile for the C43-6 dataset is shown.

b) Real space representation of the covariance matrix $\mathbf{S}(\boldsymbol{p}_{MAP})$ for the C43-6 configuration (that for the C43-3 looks qualitatively similar), showing covariance between not only adjacent but also nonadjacent points. The overall covariance decreases from the lower left to upper right because the disc brightness as represented in the visibilities decreases (on average) as a function of disc radius.

c) Draws from the prior on the brightness distribution $\mathcal{G}(\boldsymbol{I}_\nu, \mathbf{S}(\boldsymbol{p}_{MAP}))$ for both the C43-3 and C43-6 datasets. For reference the posterior mean (the frank fit for the brightness profile, Equation 30) is shown for the C43-6 dataset, separately realized using $\boldsymbol{p}_{MAP}$ estimated from the C43-6 and C43-3 dataset.

d) – f) As in (a) – (c) but using $w_{smooth} = 10^{-1}$ to generate the power spectra, which are comparatively smooth in response to this stronger hyperparameter value, resulting in reduced covariance between nonadjacent points but higher covariance between adjacent points in $\mathbf{S}(\boldsymbol{p}_{MAP})$ as visualized in (e). The reduced covariance between nonadjacent points in (e) relative to (b) results in draws from the prior in (f) that show fewer small amplitude amplitude oscillations on corresponding spatial scales. However the effect on the posterior mean frank fit is small, indicating the relative insensitivity of the recovered brightness profile to the value of the hyperparameter w_{smooth}.

are truncated (drop off) at different maximum baselines (and thus minimum spatial scales) and also show different degrees of smoothness. The covariance matrices $\mathbf{S}(\boldsymbol{p})$ in Fig. 4(b) and (e), and the draws from the prior $P(\boldsymbol{I}_\nu|\boldsymbol{p})$ in Fig. 4(c) and (f), then motivate the two effects that $\boldsymbol{p}_{MAP}$ has as a prior on $\boldsymbol{I}_\nu$.

First, because the power spectra have little power on long baselines, the prior draws are correlated on small spatial scales. The shortest length scale over which a brightness profile is correlated is controlled by the longest baseline at which the prior has significant power. In contrast, because the amplitude of the power spectra is large on short baselines, the large scale form of the brightness is free to vary, with the posterior brightness on these scales being ultimately determined by the data rather than the prior. If instead adjacent points were not correlated (if the prior were diagonal in real

space), the prior draws would not appear smooth. Fig. 4(c) and (f) additionally show the posterior means (Equation 30) – the fitted brightness profiles for the C43-6 mock data using either the C43-3 or C43-6 prior – which reflect constraints introduced by the priors. For each of the priors, the mean matches the true (input) brightness on large scales, but only shows structure on the scales allowed by the prior. For the priors generated under the C43-3 mock observations, the power on long baselines is thus too strongly damped, and the reconstructed profiles are a poor recovery of the input profile. By contrast the reconstructed profiles using the C43-6 priors recover the input profile to high accuracy. This emphasizes the importance of correctly identifying the scale on which to regularize the brightness profile so as not to damp the fit on scales where true variations exist in the visibility distribution.

Second, substructure in the power spectrum causes differences in the real space representation of the prior. The localized areas of lower power in the structured power spectra of Fig. 4(a) introduce correlations between nonadjacent radial collocation points at the corresponding spatial scales. This correlation manifests as the crosshatching in Fig. 4(b), and consequently draws of I_ν from the prior in Fig. 4(c) are more oscillatory than those in (f), for which the power spectra are smooth. However because this increased correlation corresponds to baselines at which the visibility amplitude is also small, the impact on the posterior mean brightness profiles in Fig. 4(c) is small. This need not be the case generally, but is typical of the power spectra generated by frank because these regions of low power have been determined from the visibilities. The power spectrum estimate $\tilde{I}_{\nu,\ \mathrm{C43-6}}^2$ in Fig. 4(a) and (d) further demonstrates this relative insensitivity. The comparatively large discrepancy between it and p_{MAP} in Fig. 4(d) relative to that in (a) is a result of the power spectrum estimate using $w_{\mathrm{smooth}} = 10^{-1}$ having higher amplitudes than that using $w_{\mathrm{smooth}} = 10^{-4}$ and thus providing a weaker constraint on the reconstructed brightness profile, yet this does not correspond to a less accurate brightness profile reconstruction in (f) relative to that in (c).

2.6 The reconstructed brightness profile's sensitivity to the hyperparameters

Our model for the deprojected brightness profile has five free hyperparameters: two specifying the radial points at which the profile is reconstructed, N and R_{out}; and three on the power spectrum, denoted collectively as $\boldsymbol{\beta} = \{\alpha, p_0, w_{\mathrm{smooth}}\}$, where $w_{\mathrm{smooth}} = 1/\sigma_s^2$ (analogous to the definition of the visibility weights). Definitions, default values and reasonable bounds to optionally vary these hyperparameters are summarized in **Table 1**. Here we motivate the default values and discuss the fit's sensitivity to these choices.

We have found the reconstructed brightness profile to be insensitive to R_{out} and N so long as they are sufficiently large, and also to p_0 so long as it is small relative to the power in the visibilities ($\ll 1$). The default $R_{\mathrm{out}} = 2''$ is sufficiently large for a protoplanetary disc. Alternatively if the disc radius is known, R_{out} can be set to a value somewhat (e.g., 50%) larger than this, the main constraint being that the model assumes the flux is 0 beyond R_{out}. By default N is set to 300 points to ensure that Q_{max} exceeds the maximum baseline in a dataset (which is important for fit stability).

The hyperparameters α and w_{smooth} do not appear directly in Equation 30; their effects on the brightness profile reconstruction enter entirely through their effects on p_{MAP}.

As motivated in Sec. 2.5, increasing w_{smooth} reduces structure in the fit's power spectrum, reducing correlation in the brightness profile at scales for which the power spectrum has low local amplitude, while increasing correlation between adjacent radial collocation points. However we have demonstrated in Fig. 4 that varying w_{smooth} within sensible bounds often has a negligible effect on the frank reconstructed brightness profile. As also discussed in Sec. 2.5, increasing α beyond 1.0 increases the SNR threshold below which the power spectrum falls toward p_0, damping variations in the brightness profile on scales where the visibil-
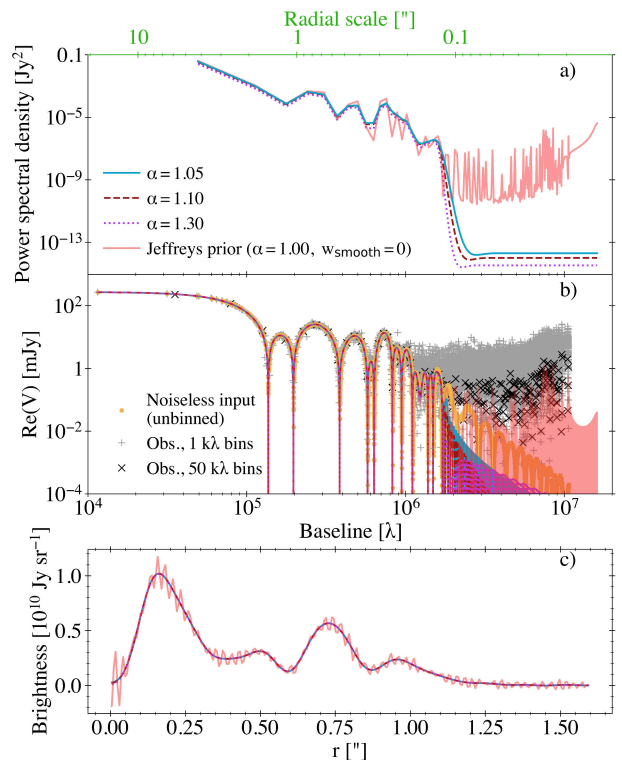
Figure 5. Effect of the hyperparameter α; effect of a Jeffreys prior
a) Maximum a posteriori power spectra p_{MAP} (under different values of the α hyperparameter) for frank fits to the mock disc in Fig. 4, here observed with the ALMA C43-6 + C43-9 configurations (beam FWHM $0.024 \times 0.030''$, Briggs=0.5; see Table 2). Also shown is p_{MAP} under a Jeffreys prior.
b) The mock observation's noisy visibilities in 1 and 50 kλ bins and the true, unbinned visibility distribution (the Hankel transform of the input profile). Also shown are the frank fits to the unbinned, noisy data and the fit under a Jeffreys prior (which forces the fit to the longest baseline, noise-dominated visibilities, imprinting oscillations on the corresponding brightness profile in (c)).
c) The fitted frank brightness profiles for the cases in (a) – (b). Varying α has a negligible effect on the fit, while the Jeffreys prior (flat in log space) allows oscillations on spatial scales corresponding to the noise-dominated visibility region (beyond ≈ 1.5 Mλ).

ity SNR is low (primarily at unconstrained scales beyond a dataset's longest baseline).

Fig. 5 examines a fit's typical sensitivity to the value of α. With $\alpha = 1.00$ (not a recommended value, shown only for pedagogy), frank will fit noise-dominated data, introducing noise into the reconstructed brightness, as in the Jeffreys prior fit in Fig. 5 that is placing significantly more power in baselines beyond ≈ 2 Mλ than exists in the noiseless input visibilities. Increasing α to 1.05 is a conservative choice, mildly damping the power on scales where the SNR is low (causing the corresponding fit in Fig. 5 to walk off the data at ≈ 1.5 Mλ). This value of 1.05 is our default choice for α because many moderate to high resolution ($\gtrsim 0.1$ Mλ) datasets for protoplanetary discs exhibit noise-dominated visibilities at their longest baselines (a consequence of sampling density

Table 1. Model hyperparameters, their default values, and reasonable ranges to optionally vary these (discussed in Sec. 2.6).

Hyperparameter	Default value (suggested range)	Referenced in
Brightness profile hyperparameters		
N, number of collocation points	300 (100 − 300)	Text following Equation 9
$R_{\rm out}$, fit's maximum radius	2″ ($\gtrsim$1.5× disc's outer edge)	Equation 4
Maximum a posteriori power spectrum hyperparameters		
α, order parameter for the power spectrum's inverse Γ distribution hyperprior	1.05 (1.00 − 1.30)	Equation 39
p_0, scale parameter for the power spectrum's inverse Γ distribution hyperprior	10^{-15} Jy^2 (> 0, $\ll 1$)	Equation 37
$w_{\rm smooth}$, strength of smoothing applied to the power spectrum	10^{-4} ($10^{-4} - 10^{-1}$)	Equation 38

Table 2. Observational quantities for all *mock* datasets. All mock observations observe at 230 GHz (1.3 mm) with 7.5 GHz continuum bandwidth.

Mock observational setup	ALMA configuration (all in Band 6)	Synthesized beam FWHM [mas] (longest baseline [10^6 λ])	Integration time [min] (sampling time [s])	Peak I [10^{10} Jy sr^{-1}] (integrated I [10^{10} Jy sr^{-1}])	Used in
Mock profiles shown in Fig. 7					
low resolution	C43-3	590×700 ($\approx$0.4)	2 (10)	1.00 (0.25 − 0.51)	Fig. 4, 6, 7, C1
moderate resolution	C43-6	130×170 ($\approx$2)	10 (10)	1.00 (0.25 − 0.51)	Fig. 2, 4, 7
high resolution	C43-6 + C43-9	24×30 ($\approx$10)	10 + 50 (10)	1.00 (0.25 − 0.51)	Fig. 5
Mock profile based on AS 209					
low noise	C43-7	86×106 ($\approx$3)	40 (10)	5.00 (0.87)	Fig. 8, 9
moderate noise	C43-7	86×106 ($\approx$3)	2 (10)	5.00 (0.87)	Fig. 9
high noise	C43-7	86×106 ($\approx$3)	1 (10)	5.00 (0.87)	Fig. 9

decreasing strongly at the most extended ALMA configurations for typical integration times). Further increasing α will more aggressively damp power on scales with low SNR, though in some cases this can lead to the fit attaining a lower effective resolution (if a significantly wider range of scales are damped). That said, comparing the $\alpha = 1.05$, 1.10 and 1.30 fits in Fig. 5 shows that varying α within sensible bounds often has an insignificant effect on the brightness profile, especially if the visibilities' SNR is dropping rapidly at long baselines. For a scenario in which varying α has a more significant impact, we will consider a disc with sub-beam features sampled at low SNR over a wide range of baselines in Fig. 9 (presented in Sec. 3.1.2).

In our tests the effect on the brightness profile of varying α and/or $w_{\rm smooth}$ is thus often trivial. Nonetheless, as a precaution *we recommend varying the hyperparameters used in a fit within sensible bounds* (see Table 1) *to assess the brightness profile's resulting sensitivity.* We have found that especially for lower resolution datasets, setting α and/or $w_{\rm smooth}$ too high can average over real, underresolved features, causing them to appear broader and shallower. Hence our default values for these hyperparameters are at the lower bound of our suggested ranges in Table 1.

2.7 Model limitations

(i) Our axisymmetric model is 1D, fitting for the azimuthal average of the visibility data at each spatial scale. In the presence of deviations from axisymmetry the model is thus biased. For mild asymmetries the effect is not severe, averaging over brightness asymmetries azimuthally. However for major asymmetries (such as a prominent spiral) and/or when (u, v) coverage results in broad gaps over a given baseline range, the model can break down. This also holds for observations in which there is more than one source.

(ii) Since a centered axisymmetric model has only real visibilities, we do not fit Im(V).

(iii) The model makes the flat sky approximation (Equation 1), which assumes that the observation's region of the sky is sufficiently small.

(iv) Before fitting for the brightness profile, the visibilities must first be deprojected and phase-centered. `frank` can optionally do this by fitting a 2D Gaussian to the visibilities, though this deprojection operation may yield an erroneous result if a disc has an appreciable vertical thickness or if limb darkening from the optically thick surface is important.

(v) The fitted brightness profile does not include primary beam correction. For sources that were observed close to the center of the primary beam, the correction is typically small and can be obtained by dividing the reconstructed brightness profile by the primary beam profile $\mathcal{A}(\theta)$, where θ is the radial angular coordinate.

(vi) Regions in the visibilities with sparse and/or sufficiently noisy sampling can cause a lack of constraint on the local spatial frequency scale, inducing oscillations in the brightness profile on the corresponding spatial scale. This can potentially mimic real structure. The model typically prevents this by damping power on scales with low SNR, but when it does occur the oscillations in the brightness profile can be diagnosed by their frequency, which corresponds to the unconstrained spatial frequency scale. *Varying the hyperparameter values for a fit as noted in Sec. 2.6 is useful to assess and potentially suppress this behavior.*

(vii) The uncertainty on the fitted brightness profile is typically underestimated. For this reason we do not show the uncertainty on `frank` brightness profiles in this work. The model framework produces an estimate of the uncertainty on the brightness profile (diag($\mathbf{D}$)), but this is not reliable because reconstructing the brightness from Fourier data is an ill-posed problem. For example if the visibility amplitude were to spike at any point beyond the data's maximum baseline, this would imprint high amplitude variations in the brightness profile on small spatial scales. Unless we know a priori (which is not generally the case) that the visibili-
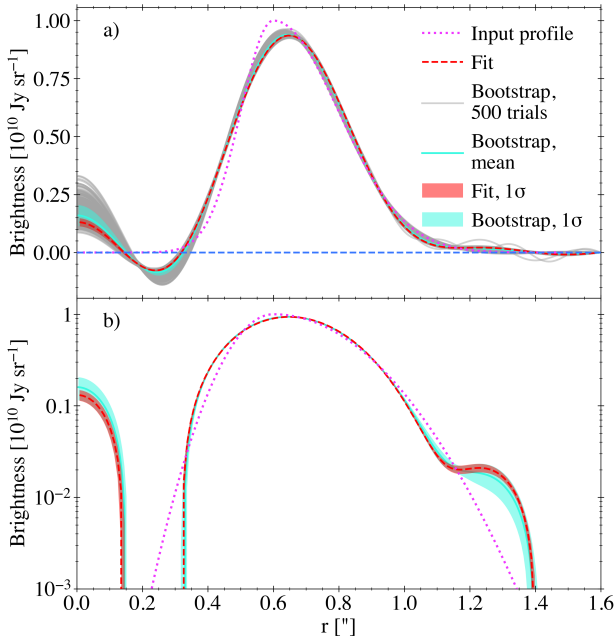
Figure 6. Underestimated model uncertainty
a) Input and reconstructed brightness profiles for a mock Gaussian ring (two joined sigmoids) observed with the ALMA C43-3 configuration (synthesized beam FWHM 0.59×0.70″, Briggs=0.5; see Table 2). The fit's 1σ uncertainty estimate, diag(**D**) estimated at the maximum a posteriori power spectrum, is shown. Additionally shown are 500 realizations of bootstrapping on this mock dataset, as well as the resulting distribution's mean and 1σ uncertainty. Both the frank fit's uncertainty and the bootstrap uncertainty are clearly underestimated as discussed in Sec. 2.7.
b) As in (a) on a logarithmic scale.

ties are decreasing sufficiently rapidly with increasing baseline, the uncertainty is therefore formally infinite. While it is reasonable to assume that for real disc brightness profiles the visibilities do decrease rapidly at long baseline, it is not straightforward to generically extrapolate the slope of this decline beyond a dataset's longest baseline; a robust error estimate is thus difficult to obtain.

Fig. 6 shows the frank uncertainty estimate, diag(**D**) estimated at the maximum a posteriori power spectrum, for the mock Gaussian ring presented in Sec. 3.1. This confidence interval approximately represents the fit's statistical uncertainty (that due to the uncertainty on the observed baselines), which is correct if the visibility weights are an accurate representation of the pointwise visibility uncertainty. But the confidence interval does not capture the fit's ill-defined systematic uncertainty (that due to sparse sampling in the (u, v) plane). A bootstrap on the visibilities in Fig. 6 also fails to yield a reasonable estimate of the systematic uncertainty. Though this does confirm that diag(**D**) is a reasonable estimate of the statistical uncertainty. *We therefore urge caution before using uncertainty estimates to interpret the significance of the result in the recovered profile.*

To test whether including the uncertainty on p has a significant effect on the final uncertainty of the reconstructed brightness, we used the estimate for the uncertainty on p given by Oppermann et al. (2013). We translated the effect of this on the brightness profile by Taylor expanding $P(\boldsymbol{V}, \boldsymbol{p})$

about its maximum to make a Gaussian approximation to $P(\boldsymbol{V}, \boldsymbol{p})$. Using this approximation, we then drew samples and compared the variance of the reconstructed brightness at each radial collocation point to that estimated by diag(**D**). Except for an uninformative Jeffreys prior ($\alpha = 1.0$ and $w_{\text{smooth}} = 0$), the effect was negligible.

(viii) The fitted brightness profile can have negative regions corresponding to spatial scales un- or underconstrained by the visibilities. There is an argument for choosing a fitting strategy that enforces the solution be nonnegative (as in Junklewitz et al. 2016), and we have investigated the effect of negative fit regions by finding the most probable nonnegative intensity profile given $\boldsymbol{p} = \boldsymbol{p}_{\text{MAP}}$. The effect on the recovered brightness profile is localized to the regions of negative flux, with otherwise minor differences. We explore a pedagogical nonnegative fit in Appendix C.

2.8   Code performance

The code's computation time is dominated by two components, constructing the matrix **M** and iterating the fit. The construction of **M** is only done once at the start of the fit and has computational cost $O(N^2_{\text{collocation points}} N_{\text{visibilities}})$, while solving the linear systems in each iteration scales as $O(N^3_{\text{collocation points}}) \sim 100^3$.

To limit memory requirement, the matrix **M** is assembled in blocks, avoiding the need to hold the $N_{\text{collocation points}} \times N_{\text{visibilities}}$ matrix **H**($\boldsymbol{q}$) in memory. Typical computational requirements were estimated using real datasets: with 10^4 visibilities and 200 collocation points the fit took 10 s and used $\approx$100 MB, while with 10^6 visibilities it took 40 s and used $\approx$200 MB. These tests were conducted on a 2017 MacBook Pro with a 7th generation Intel Core i5 processor (7360U) running at 2.3 GHz with 8 GB RAM.

3   DEMONSTRATION & ANALYSIS

3.1   Demonstration on mock observations

3.1.1   *Fits can attain sub-beam resolution*

To demonstrate frank's fitting approach and characterize its performance, we begin in **Fig. 7** with a series of mock observations for discs of archetypal smooth and sharp structure at both well resolved and underresolved scales. We generate these mock datasets using the default fit hyperparameters in Table 1 and the observational setups in **Table 2**.

Fig. 7(a) compares the brightness reconstructed by frank to the CLEAN image-extracted profile for a Gaussian disc centered at 0. The Gaussian's width is resolved by the CLEAN beam for the C43-6 ALMA configuration, yet importantly the CLEAN profile is slightly too broad and shallow as a result of beam convolution[5]. By comparison frank recovers the profile to <1% RMS error (note this is

---

[5] As a simple illustration of this effect, noiseless Gaussians with FWHM 25, 50, 75, and 100 mas would respectively be broadened by convolution with a FWHM 50×50 mas Gaussian beam to FWHM 56, 71, 90, and 112 mas (a factor of 2.24, 1.42, 1.20, and 1.12 increase), and their amplitudes would be reduced to 45, 70, 83 and 89% their true values.
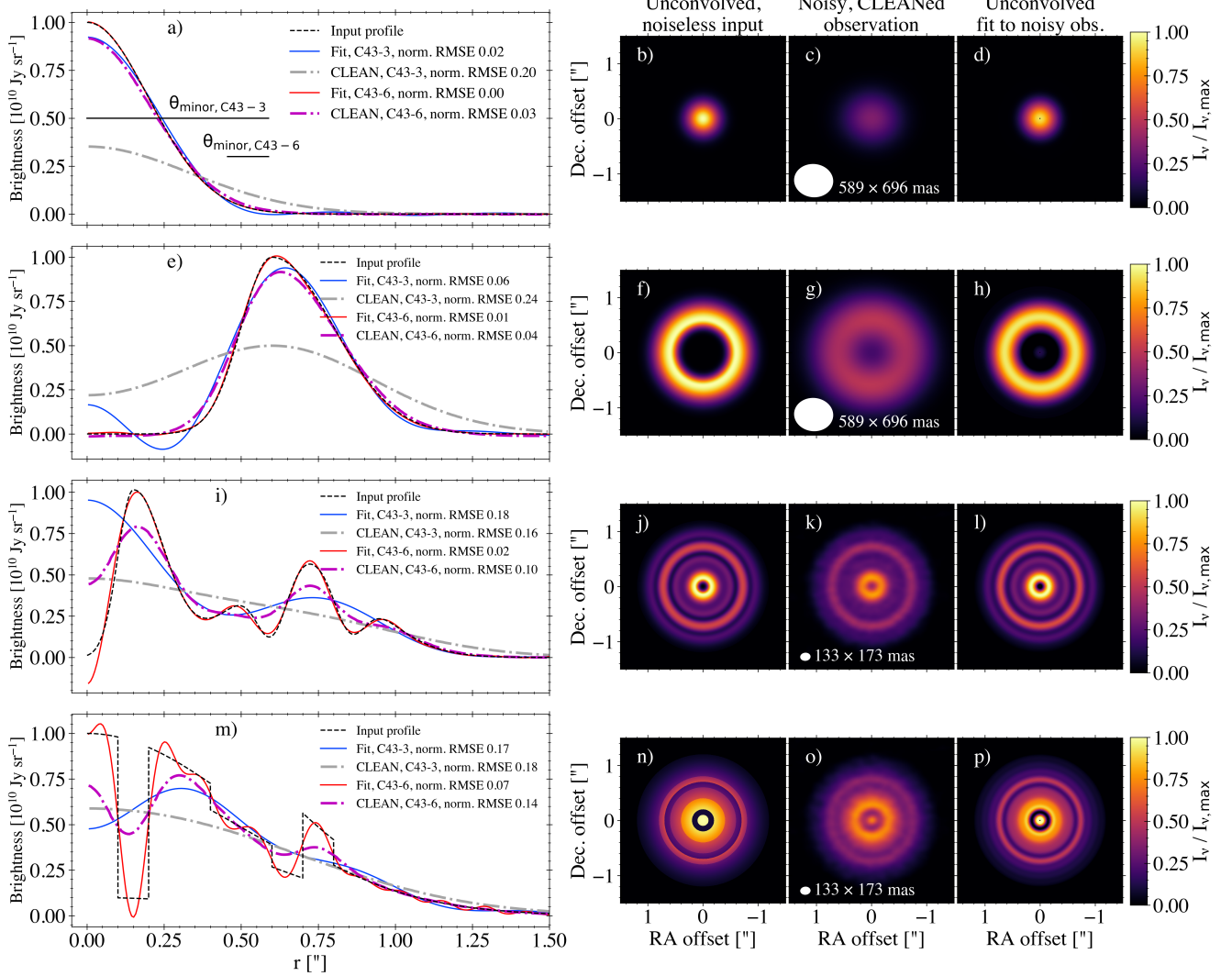
Figure 7. Performance on different disc morphologies using mock observations

a) Input and reconstructed brightness profiles for a mock Gaussian disc with FWHM 0.2″, observed separately with the ALMA C43-3 (synthesized beam FWHM 0.59×0.70″, Briggs=0.5) and C43-6 (beam FWHM 0.13×0.17″, Briggs=0.5, factor of 5 longer integration time than C43-3; see Table 2) configurations. The beams' minor axes are shown for reference. The `frank` fit to each of the 2 mock observations is shown, as are CLEAN image-extracted profiles. The normalized RMS error of each profile is given.

b) Input profile swept over 2π, noiseless and at infinite resolution.

c) Mock CLEAN image (C43-3 + noise), with brightness normalized to (b).

d) Image of the `frank` fit to the noisy C43-3 mock observation, *not* convolved with the beam, with brightness normalized to (b).

e) − h) As in (a) − (d) but for a Gaussian ring (two joined sigmoids) with width ≈ 0.4″.

i) − l) As in (a) − (d) but for a more complicated, smooth-featured disc. The feature widths vary from 0.10−0.15″. The images correspond to the C43-6 observation.

m) − p) As in (i) − (l) but for a disc with step features, the narrowest of which is 0.1″.

the standard definition of RMS error, unrelated to the RMS noise in a CLEAN image). For the same disc observed at the lower resolution C43-3 configuration, the beam's minor axis underresolves the Gaussian by a factor of 2.3, and the effect of beam convolution on the CLEAN profile is exacerbated. The `frank` fit by comparison retains high accuracy, comparable to the CLEAN profile from the C43-6 observation. Fig. 7(b) − (d) show the 2D images for the C43-3 case, with the `frank` image recovering the input image from the noisy, low resolution observation. This simple Gaussian case illustrates that because the model does not at any stage

require beam convolution, `frank`'s achievable resolution is sub-beam.

Fig. 7(e) − (h) next consider a slightly more complex disc, a Gaussian ring. Here again the `frank` fit to the C43-3 observation achieves a similar accuracy to the CLEAN profile extracted from the higher resolution C43-6 dataset. This is despite the C43-3 beam's minor axis underresolving the Gaussian by a factor of ≈1.5. The `frank` profile reconstructed from the C43-3 dataset does however misidentify the Gaussian's centroid and also shows a region of negative brightness allowed by the model as detailed in Sec. 2.7. We

discuss a fit with enforced brightness positivity for this disc in Appendix C.

As well as reproducing simple profiles, `frank` adapts effectively to more complicated discs such as that in Fig. 7(i) – (l). Here the C43-3 beam's minor axis is a factor of >3 broader than the profile's widest feature, and `frank` is unable to reconstruct the profile accurately. However it does discern that there are two well-separated peaks, while the CLEAN profile does not show any substructure. Increasing the resolution to C43-6, `frank` retains its resolving power advantage, recovering the input brightness profile to high accuracy.

To strain the model, Fig. 7(m) – (p) introduces perfectly sharp-edged features (step functions) to be recovered. `frank` fits this disc to reasonable fidelity in the C43-6 case but does show some oscillations. These are a consequence of Gibbs phenomenon, which arises when representing an infinitely sharp feature in Fourier space. By comparison the CLEAN profiles at both resolutions do not show these oscillations, but do smear the disc features over the beam. `frank`'s ability to recover arbitrarily sharp features with comparatively low error demonstrates utility in more accurately recovering a disc's often steep outer edge and its peak flux.

Together these mock discs show `frank`'s ability to fit smooth and sharp, partially and well-resolved, faint and bright features, at sub-beam resolution. In practice this can enable similar fit resolution to a CLEAN profile obtained with a more extended array configuration (e.g., the `frank` fit to the C43-3 data in Fig. 7(a) is comparable to the CLEAN profile for the C43-6 data). The `frank` fits for all 8 mock datasets shown in Fig. 7 recover the discs' total flux to within a mean 0.8% (standard deviation 0.2%), compared to a mean 1.6% (standard deviation 1.6%) for the CLEAN profiles. This error in total flux recovery increases as a disc's features become increasingly sub-beam, an effect that is more severe for CLEAN than `frank`. All `frank` fits shown here are negligibly sensitive to the choice of hyperparameter values within the suggested ranges listed in Table 1. As discussed in Sec. 2.8 these fits – and all others shown in this work – are performed in ≲1 min, and the computation speed is independent of the complexity of disc substructures; `frank` *fits simple and complicated disc profiles equally fast.*

To illustrate how `frank` is attaining a sub-beam fit resolution, **Fig. 8** characterizes its performance using a mock profile based on observational data rather than a simple functional form. We use the CLEAN image-extracted fit to the real DSHARP observations (C40-5/8/9 + archival short and moderate baseline datasets) of AS 209 (Andrews et al. 2018) as the input profile to be recovered from mock observations. This profile was obtained in that work with a beam of FWHM 36×38 mas; we generate the mock data at a factor of ≈2.6 worse resolution using the C43-7 configuration (beam FWHM 86 × 106 mas, see Table 2). Much of the structure in the profile is thus sub-beam in the mock observation. The mock visibilities' SNR as a function of baseline is similar to the real dataset.

`frank` accurately recovers the input profile's sub-beam features to within 1% RMS error for the noiseless case in Fig. 8(a) – (b). It shows minor difference between the fit under the default hyperparameter values and under values at the other extrema of our suggested range (Table 1) as shown in Fig. 8(c), despite the difference in prior structure

in Fig. 8(f). The brightness profile recovery's high fidelity is a result of the model fitting the visibility distribution in Fig. 8(d) – (e) to high accuracy. As in previous cases the CLEAN profile extracted from the C34-7 mock image underresolves these features. The `frank` fit recovers the mock disc's total flux to within 0.6%, while the analogous CLEAN error is 30.4%. The nontrivial CLEAN error is primarily a result of the mock observations containing no baselines shorter than those in the ALMA C43-7 configuration; i.e., the nominal maximum recoverable scale for C43-7, Band 6 is 1.12″ (Remijan et al. 2019), while the disc extends to 1.25″. By comparison, `frank` is able to extrapolate the fit accurately to short baselines despite this lack of data (though this behavior may not always be robust). In Fig. 8(f) the maximum a posteriori power spectrum p_{MAP} has low power at the unconstrained spatial frequencies beyond the data's longest baseline, preventing large amplitude oscillations in the reconstructed brightness profile at equivalent spatial scales (these oscillations, damped, are seen in the residuals of Fig. 8(c)). In Fig. 8(g) – (i) the 2D images of the input model, mock CLEANed observation and fit demonstrate the amount of structure `frank` is recovering that is smeared out by the beam.

Beam convolution is the primary difference in achievable resolution between `frank` *and CLEAN.* Convolving either `frank` fit in Fig. 8(a) – (b) with the 2D CLEAN beam yields a profile that is similar to the CLEAN profile. The CLEAN beam size does depend on the visibility weighting, with the choice affecting the resulting CLEAN profile. For the disc in Fig. 8, setting the Briggs robust parameter to −2.0 (approximating uniform weighting), 0.5 and 2.0 (natural weighting) sets the beam FWHM as 0.08″×0.10″, 0.09″× 0.10″ and 0.10″ × 0.12″. However the resulting profiles vary only slightly, with the RMSE changing at most by 3%, and each profile still underresolves the disc features relative to `frank`. This highlights that *`frank` can recover disc features underresolved even by uniform weighting, while retaining high sensitivity.*

3.1.2 *Baseline-dependent signal-to-noise determines the achievable fit resolution*

We next characterize the model's SNR sensitivity by decrementing the integration time for mock observations of the input profile from Fig. 8. This increasingly degrades the (u, v) coverage, in turn worsening the data's effective SNR at a given baseline. **Fig. 9**(a) – (b) first show the method's intrinsic capability by fitting the *noiseless* data, with the (u, v) coverage determined by the mock observation's C43-7 configuration and integration time. The fits using both values for the α hyperparameter accurately match the full visibility distribution, the reconstructed brightness profiles showing <1% RMS error.

Then fitting the same baseline distribution but with the mock observation's noise included in Fig. 9(c) – (d), both fits remain able to recover the noiseless visibilities to high accuracy. However as the data's SNR worsens beyond ≈1.8 Mλ the fits do show some error, including tending toward 0 prior to the longest baselines. These errors manifest in the recovered brightness profiles in Fig. 9(c) as a slightly less accurate recovery of the peak brightness and of the innermost gap's depth. The fit using the stronger $\alpha = 1.30$ is slightly
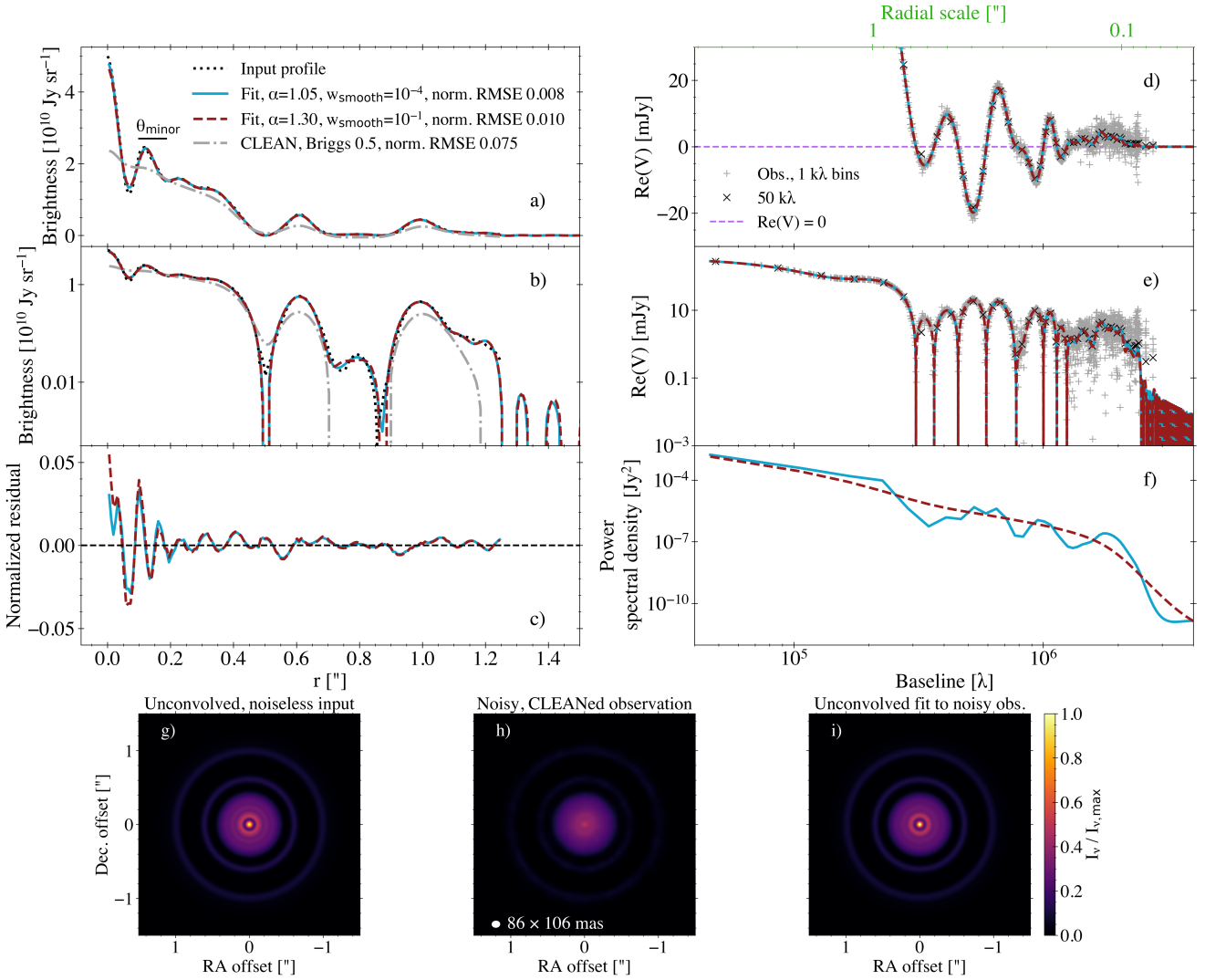
Figure 8. Fit methodology using mock observations
a) Input and reconstructed brightness profiles under different hyperparameter values for a mock disc that uses as input the CLEAN image-extracted fit to the real DSHARP AS 209 observations (ALMA configurations C40-5/8/9, synthesized beam FWHM 36 × 38 mas; Andrews et al. 2018). The mock observation uses only the C43-7 configuration (beam FWHM 86 × 106 mas, Briggs=0.5; see Table 2). This beam's minor axis is shown for reference. The CLEAN image-extracted profile for the mock C43-7 observation is also shown, with the normalized RMS error of all fits given.
b) As in (a) on a logarithmic y-scale. The fit's oscillations beyond the disc's outer edge indicate the model's noise floor.
c) Normalized residual for the `frank` fits under different hyperparameter values.
d) Visibilities for the mock C43-7 observation in (a). The brightness profile fits in (a) are the discrete Hankel transforms of these visibility domain fits.
e) As in (d) on a logarithmic y-scale.
f) The fit's maximum a posteriori power spectrum p_{MAP} for different hyperparameter values. These power spectra are used as the priors on the brightness profile reconstructions in (a).
g) Input profile swept over 2π, noiseless and at infinite resolution.
h) Mock CLEAN image (C43-7 + noise), with brightness normalized to (g).
i) Image of the `frank` fit to the noisy C43-7 mock observation, *not* convolved with the beam, with brightness normalized to (g).

less accurate at the longest baselines because the mock observations do not have sufficiently high SNR there for the model to fit them with $\alpha = 1.30$. However the differences in the reconstructed profile are minor for this case.

As the integration time is further decremented in Fig. 9(e) − (f), the fits' fidelity is degraded primarily beyond ≈1.2 Mλ as the visibility distribution becomes more sparse.

By consequence the highest resolution structural information – the depths and centroids of the narrowest features in the input brightness profile, including the peak brightness – are only partially recovered. The effect is more severe for the $\alpha = 1.30$ case because the data beyond 1.2 Mλ have low effective SNR, and as α increases, the SNR threshold below which `frank` does not fit the visibilities increases. Since these
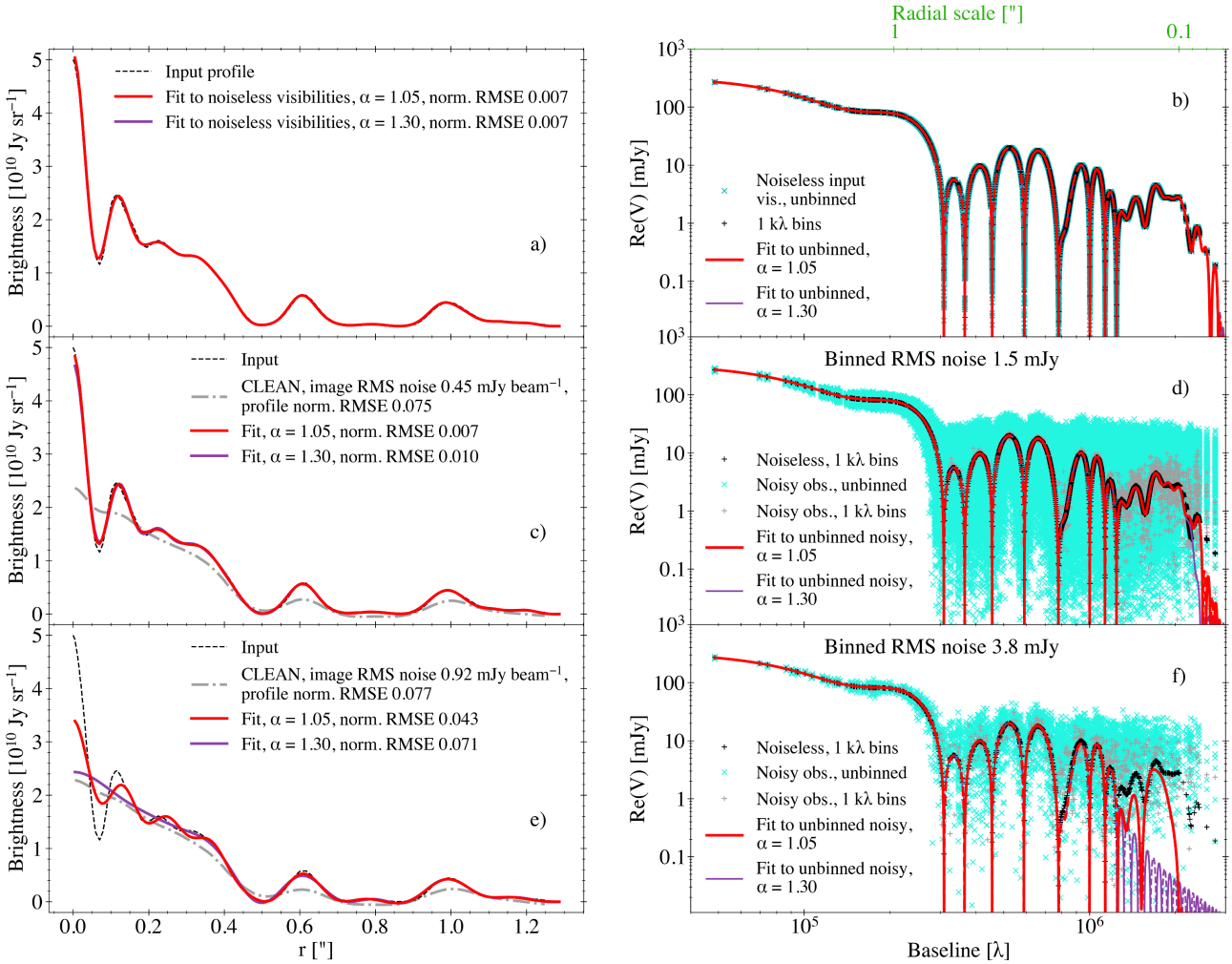
Figure 9. Fit sensitivity to baseline-dependent SNR using mock observations
a) The input profile from Fig. 8 is used to generate noiseless mock observations, with `frank` fits to these noiseless data shown and their normalized RMS errors given under two values of the α hyperparameter (both fits use $w_{smooth} = 10^{-4}$).
b) Noiseless visibilities corresponding to the input profile in (a). `frank` fits the *unbinned* visibilities shown. Also shown are the data in 1 kλ bins. The data (and the fits') negative regions are denoted by the fit lines' dashed sections. The $\alpha = 1.30$ fit is behind the $\alpha = 1.05$ fit.
c) – d) As in (a) – (b) but with noise added to the mock observation. The fits shown are to the unbinned, noisy data. The CLEAN image-extracted profile is shown for comparison. The apparently larger scatter in the observations in (d) (compared with (f)) is due to the larger number of data points rather than higher intrinsic error.
e) – f) As in (c) – (d) but with a higher RMS noise due to shorter integration time. The binned RMS noise (1 kλ bins) does not increase dramatically because the number of empty bins has also increased.

data clearly contain meaningful information, $\alpha = 1.05$ is the more sensible choice.

In contrast to `frank`'s SNR sensitivity, the CLEAN image-extracted profiles in Fig. 9(c) and (e) vary marginally when the visibilities' SNR is decreased. This is because convolution with the C43-7 beam, rather than the data's SNR, is primarily limiting the CLEAN image resolution. While `frank`'s resolving power is sensitive to variations in the baseline dependence of the data's SNR, the CLEAN profile is largely determined by the pure baseline distribution. This entails that *improving an observation's SNR via the on-source integration time can significantly enhance the resolving capability of* `frank`, *while it may make little difference for CLEAN.*

3.2 Demonstration on real observations

Mock datasets are useful to test and characterize `frank`'s performance, though real data have more complex noise properties to which the model must also be well suited.

3.2.1 *Sub-beam fits: Characterizing fine structure in real, high resolution observations*

To this end **Fig. 10** demonstrates `frank`'s performance with real data, fitting the high resolution DSHARP observations

of AS 209 (synthesized beam FWHM 36×38 mas ≈ 5 AU at 121 pc)[6]. Fig. 10(a) – (b) show the `frank` fit (using the default hyperparameter values) and the CLEAN image-extracted profile from Andrews et al. (2018). `frank` recovers additional substructure and higher amplitude features in the inner disc, a higher peak brightness, and slightly narrower rings in the outer disc. These results are consistent with those using mock observations in Sec. 3.1. That the `frank` fit finds a negative innermost gap indicates that the innermost disc's features are not fully resolved; forcing the fit to be positive results in the innermost gap having zero flux and alters the profile's integral by 4%, but otherwise has a negligible effect (see Appendix C for a discussion of nonnegative fits). The `frank` fit agrees with the CLEAN profile in finding that the gaps centered at ≈ 0.5 and $0.8''$ are not empty[7]. However the fine structure in these gaps remains uncertain because their brightness is near the fit's noise floor (which can be approximated by the amplitude of the oscillations beyond $\approx 1.3''$).

The `frank` fit convolved with the synthesized beam is also shown in Fig. 10(a) – (b), its similarity to the CLEAN profile giving credence to `frank` correctly finding real higher resolution structural information. The sub-beam resolution of the reconstructed brightness profile is enabled by `frank` accurately fitting the visibilities in Fig. 10(c) – (d) out to ≈ 4.5 Mλ. This is equivalent to an angular scale $\sim 1/(4.5$ M$\lambda) = 46$ mas $= 5.5$ AU, which is an indication of the data's effective resolution beyond which the visibilities are noise-dominated. In Fig. 10(e) a histogram of the binned visibilities commensurately decreases sharply in counts beyond ≈ 4.5 Mλ. Fitting longer baselines with our current modeling approach only imprints noise on the brightness profile as discussed in Sec. 3.2.2.

Although the CLEAN beam has a FWHM 36×38 mas that is less than the 46 mas equivalent of the `frank` visibility fit, these two values are not directly comparable. A more direct comparison can be made in Fourier space; the DHT of the CLEAN profile in Fig. 10(c) – (d) demonstrates how convolution with the CLEAN beam effectively acts as a lowpass filter, downweighting the contribution to the CLEAN brightness profile from visibilities beyond ≈ 2.4 M$\lambda \Leftrightarrow 86$ mas $= 10.4$ AU. This is the baseline at which the DHT of the CLEAN profile begins to show discrepancies with the observed visibilities. Convolution with the beam thus suppresses features in the profile on spatial scales smaller than 10.4 AU, causing them to appear broader and shallower than in the `frank` reconstruction, which fits the visibilities out to ≈ 4.5 M$\lambda \Leftrightarrow 46$ mas $= 5.5$ AU.

The DHT of the CLEAN profile more generally shows a less accurate agreement with the visibilities than `frank` beyond ≈ 1.5 Mλ, motivating that the `frank` brightness profile is correctly identifying high resolution structure. For comparison the *parametric* visibility domain fit in Guzmán et al. (2018) uses the CLEAN image to motivate an 8 Gaussian functional form for the brightness profile, shown in Fig. 10(a) – (b). The Fourier transform of this parametric form in Fig. 10(c) – (d) accurately fits the visibilities out to ≈ 2.5 Mλ but begins to show discrepancies at longer baselines. The brightness profile has less structure in the inner disc and lower amplitude features than the `frank` fit. Because the `frank` fit accurately traces the data out to ≈ 4.5 Mλ, its higher achieved resolution finds the disc features to be narrower and higher in amplitude (though the Guzmán et al. 2018 fit is positive everywhere). This is an example of the comparative advantage of a nonparametric form to fit a complicated visibility distribution.

Fig. 10(f) – (i) compare the CLEAN image with the image of the *unconvolved and deprojected* `frank` fit, the fit convolved with the beam, and the convolved residual image of the fit. The latter demonstrates an additional use case of the model, identifying and isolating small deviations from axisymmetry (for large deviations from axisymmetry, our axisymmetric fits may not be reliable). Here the $5 - 10\%$ deviations in the brightness around each ring may potentially be explained by the gaps and rings being produced by the combination of a planet and a low viscosity as suggested in (Fedele et al. 2018). In such a case these asymmetries may be expected as a result of the low viscosity (Hallam & Paardekooper 2020).

Varying the fit hyperparameters α and w_{smooth} within sensible bounds has negligible effect on the `frank` brightness profile. **Fig. 11** compares the fit from Fig. 10 using the default values ($\alpha = 1.05, w_{\mathrm{smooth}} = 10^{-4}$) with a fit that more strongly damps low SNR data ($\alpha = 1.30, w_{\mathrm{smooth}} = 10^{-1}$). The latter smooths the power spectrum appreciably relative to the default choice in Fig. 11(a), yet the effects on the visibility fit and the reconstructed brightness in Fig. 11(b) – (c) are negligible; the fit is robust to how precisely the prior weights the longest (i.e., noisiest) baselines. Note that although the priors (power spectra) shown in Fig. 11(a) do not extend to the shortest baselines, this does not significantly impact the fit. This is because the maximum recoverable scale of the observations is much larger than the size of the disc, and `frank` recovers the brightness accurately so long as R_{out} is larger than the disc size.

Although convolving the `frank` profile with the CLEAN beam generally results in brightness profiles that are similar to CLEAN-extracted profiles, the converse is not true. In **Fig. 12** we compare a profile extracted from the raw (unconvolved) CLEAN point source model (*.model* image); a profile extracted from the point source model convolved with the CLEAN beam (without the addition of the residual dirty map, the *.residual* image); and a profile extracted from the final CLEAN *.image* image. One may expect that the fit resolution achieved in the CLEAN point source model profile is similar to that in the `frank` profile. While the disc's innermost gap is narrower in the point source model profile relative to the final CLEAN image profile, the point source model profile does not show as much structure in the inner disc as the `frank` fit. Moreover the noise in the point

---

[6] We downloaded the AS 209 self-calibrated and multiconfiguration combined continuum measurement sets from https://bulk.cv.nrao.edu/almadata/lp/DSHARP. Before extracting the visibilities using the `export_uvtable` function of the `uvplot` package (Tazzari 2017), we applied channel averaging (to obtain 1 channel per spectral window) and time averaging (30 sec) to all spectral windows and multi-configuration datasets in the original MS table.
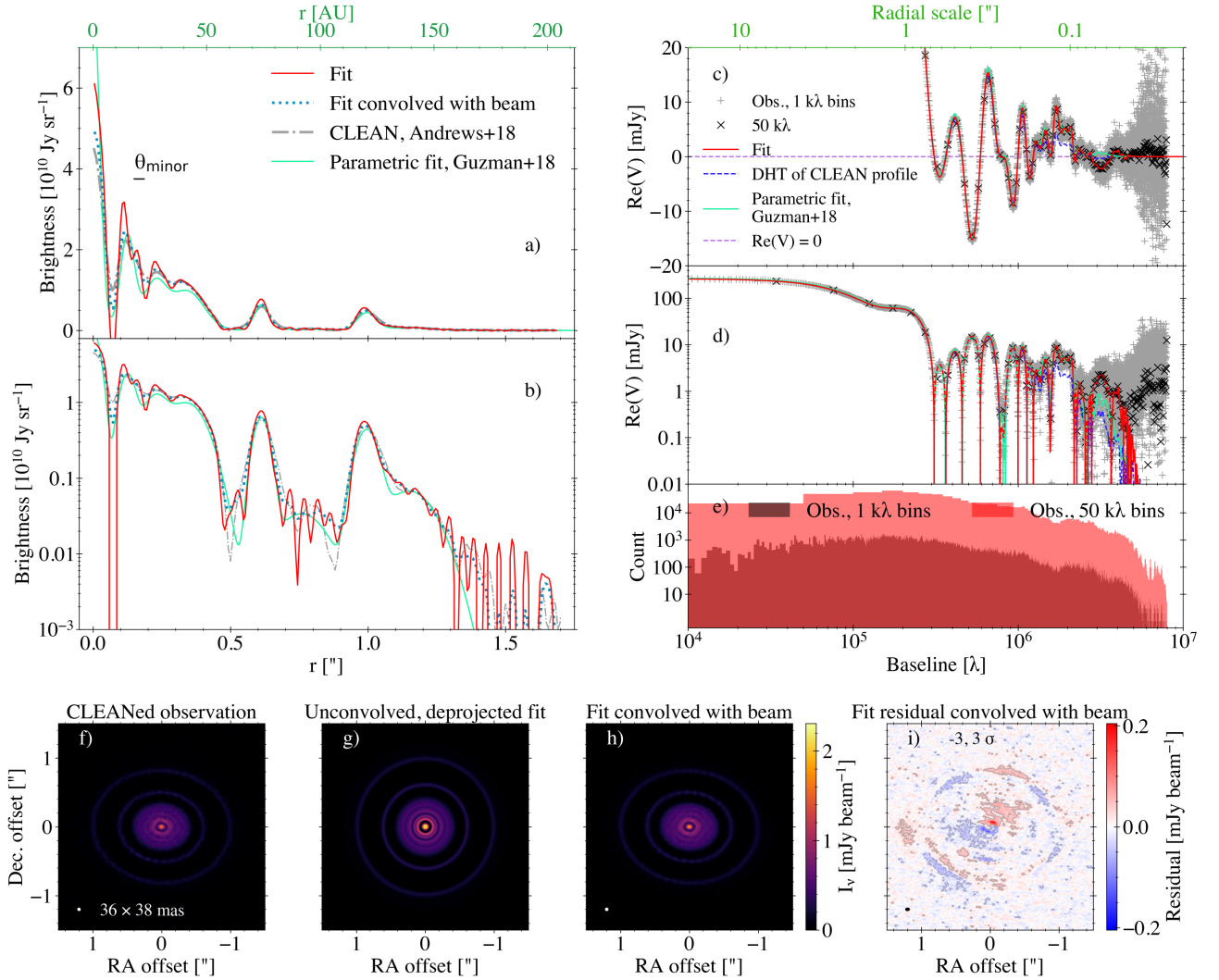
[7] Although the SNR in the gaps is low, this noise is dominated by the uncertainty on scales $\lesssim 0.05''$; it is the lower uncertainty on scales of the approximate gap widths, $0.1''$ and $0.3''$, that suggests the average flux in the gaps is nonzero.

**Figure 10. Fit to real, high resolution observations: AS 209**

a) Reconstructed brightness profile for the real, high resolution (beam FWHM 36×38 mas) DSHARP observations of AS 209 in Andrews et al. (2018). The CLEAN image-extracted profile from that work and the beam's minor axis are shown. Consistent with `frank` demonstrations on mock data, the model achieves a higher fit resolution (narrower, higher amplitude features) than the CLEAN technique. The presence of additional structure in the inner disc relative to the CLEAN image-extracted profile is robust to variations of the model. Also shown is the `frank` brightness profile convolved with the observations' 2D synthesized beam and the brightness profile for the parametric visibility domain fit in Guzmán et al. (2018).

b) As in (a) on a logarithmic y-scale. The outermost radii show the fit's noise floor.

c) Zoom on the region around 0 of the observed DSHARP visibilities in 1 and 50 kλ bins. The `frank` fit, the discrete Hankel transform of the CLEAN profile in (a), and the parametric fit from Guzmán et al. (2018) corresponding to the profile in (a) are shown. The strong increase in scatter at $\gtrsim 4.5$ Mλ sets the data's *effective* resolution, beyond which `frank` does not attempt to fit the noise-dominated visibilities.

d) As in (c) on a logarithmic y-scale. The data (and the `frank` fit's) negative regions are denoted by the `frank` fit line's dashed sections.

e) Histogram of (binned) visibility counts, showing the strong decrease beyond ≈ 4.5 Mλ responsible for the increased scatter in (c) − (d).

f) CLEANed observation.

g) Image of the *unconvolved and deprojected* `frank` reconstruction.

h) Image of the `frank` reconstruction reprojected and convolved with the observations' synthesized beam.

i) Residual map of the `frank` reconstruction (the fit's residual visibilities imaged with CLEAN), showing evidence of non-axisymmetric structure (contour $\sigma = 10$ μJy beam^{-1}, peak residual 12σ).
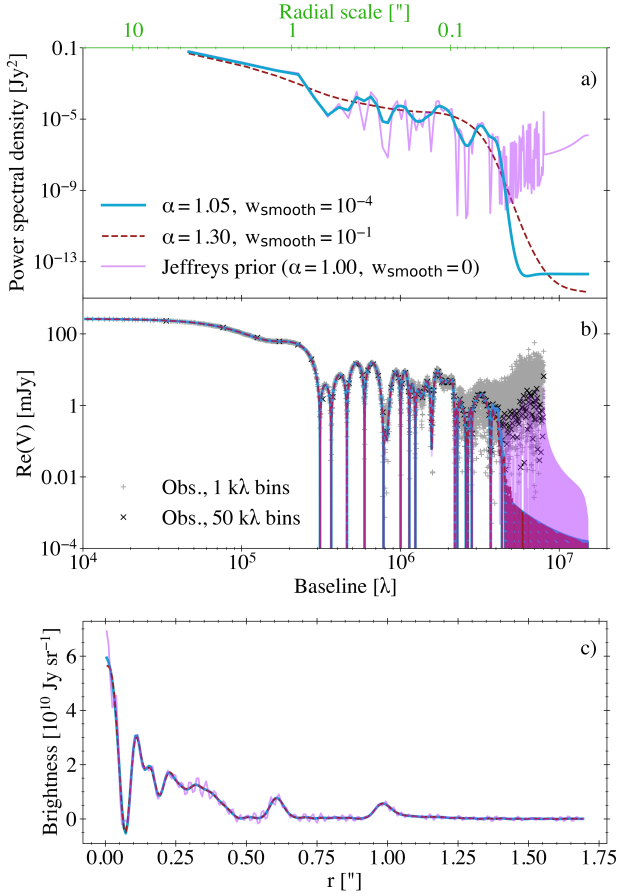
Figure 11. Fit sensitivity to hyperparameters: AS 209
a) Maximum a posteriori power spectra of the `frank` fit under the default hyperparameter values ($\alpha = 1.05$, $w_{smooth} = 10^{-4}$) and stronger values ($\alpha = 1.30$, $w_{smooth} = 10^{-1}$). The structural differences between these spectra show negligible effect on the visibility fits in (b) and fitted profiles in (c). The power spectra are the priors placed on the respective brightness profile reconstructions. The maximum a posteriori power spectrum under a Jeffreys prior ($\alpha = 1.00$, $w_{smooth} = 0$) is also shown, which forces the model to fit the longest baseline, noise-dominated visibilities in (b), imprinting oscillations on the corresponding brightness profile in (c).
b) Observed visibilities in 1 and 50 kλ bins; the `frank` fit under the default hyperparameter values and under a stronger choice, showing minor difference; the fit under a Jeffreys prior.
c) The fitted `frank` brightness profile in Fig. 10(a) – (b) under the default hyperparameter values, as well as fits under the stronger choice and Jeffreys prior.

source model profile (even when binning) can make inference on disc feature widths and amplitudes nontrivial. We have applied this same analysis to multiple real and mock datasets at various antenna configurations, finding in general that while the resolution difference between profiles extracted from the *.model* image and *.image* image is starker in lower resolution data (because the beam is larger and so convolution with it has a stronger effect), the point source model profile's noise also worsens for lower resolution data. We have confirmed this behavior with a CLEAN gain parameter as low as 10^{-3}.
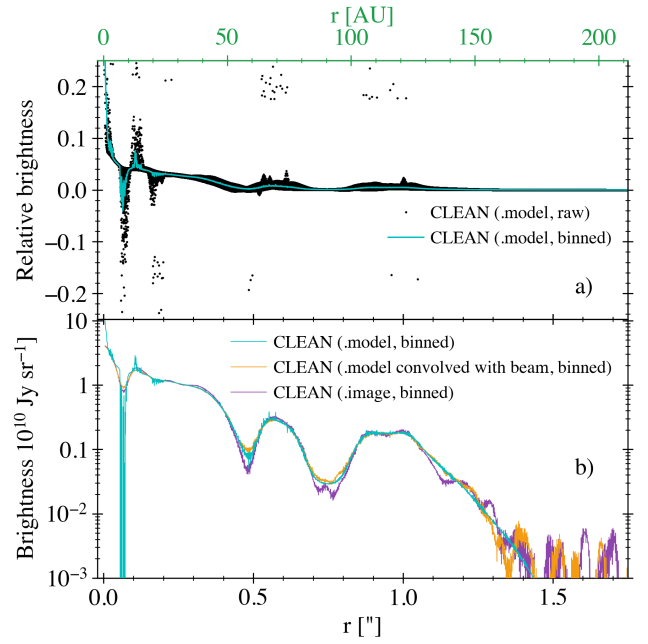
Figure 12. CLEAN model profile extraction: AS 209
a) Pixel flux as a function of disc radius, extracted from the CLEAN point source model (*.model* image) for the DSHARP AS 209 dataset. A binned average (1 mas bins) of these data is also shown to estimate the radial profile underlying the noisy pixel flux.
b) Binned average from (a), compared with a profile extracted (using the same binning) from the point source model convolved with the CLEAN beam (*.image - .residual*), and a profile extracted from the final CLEAN image (*.image*). The lack of prominent additional disc features in the *.model* profile indicates a comparable resolution to the *.image* profile.

3.2.2 A dataset's effective resolution can be much less than the longest baseline

To further demonstrate that the longest baselines in datasets are often noise-dominated, we find that both the `frank` and CLEAN brightness profiles change negligibly if the AS 209 dataset, which extends to 8 Mλ ⇔ 26 mas = 3.1 AU, is truncated at 5 Mλ ⇔ 41 mas = 5.0 AU *prior to* the visibilities being fit. This suggests that the data's effective resolution as seen by both CLEAN and `frank` is $\lesssim 5$ Mλ, or at $\lesssim 63\%$ of the baseline distribution. In Fig. 10(c) the visibilities begin oscillating rapidly about 0 beyond ≈5 Mλ, indicative of noise dominating the signal. Fitting longer baselines only imposes noise on the reconstructed brightness profile as shown by the Jeffreys prior fit in Fig. 11 (this behavior is consistent with that found using mock data in Fig. 5). The Jeffreys prior causes the model to fit the full baseline distribution, including the noise-dominated region beyond ≈ 4.5 Mλ. This imposes oscillations on the brightness profile, demonstrating that these baselines in the AS 209 dataset are noise-dominated. The fit with the default hyperparameter values is thus (at least approximately) correctly identifying where the data become noise-dominated and accordingly justified in not fitting beyond ≈4.5 Mλ. This is not to say that an alternative fitting approach could not obtain useful information on disc structures from these noise-dominated data.

Importantly though *it is common for an appreciable fraction of the long baselines in real datasets to be noise-dominated.* This occurs in high as well as low – moderate resolution observations and suggests the integration time at the most extended baselines is often too short to yield informative visibility measurements for CLEAN or `frank`. This 'effective resolution' is predominantly a result of sampling density dropping appreciably at the longest baselines in many real datasets (while the longest baselines also typically exhibit higher phase noise, self-calibration can often alleviate this).

3.2.3 *Sub-beam fits: Identifying underresolved features in real, moderate resolution observations*

While `frank` has utility in more accurately characterizing disc substructure in high resolution observations, it is also effective in identifying sub-beam structure in low – moderate resolution observations. **Fig. 13** compares the `frank` and CLEAN fits to the DSHARP AS 209 dataset from Fig. 10 with fits to lower resolution (longest baseline 2 Mλ) observations of the same disc (Fedele et al. 2018). Relative to CLEAN, the `frank` fit to the lower resolution dataset is more accurately identifying a number of features in the brightness distribution (the centroids and widths of the outer disc's rings, depths of the adjacent gaps, and the disc's outer edge) as confirmed by the higher resolution DSHARP observations. The `frank` fit to the lower resolution data also shows reasonable agreement with the CLEAN fit to the high resolution DSHARP data in the outer disc. This is due to both profiles fitting the visibilities out to a similar maximum baseline (the `frank` fit to the lower resolution dataset extends to ≈2.1 Mλ, and recall the DHT of the DSHARP CLEAN beam's FWHM is ≈2.4 Mλ). `frank`'s sub-beam resolving capability is thus not limited to high resolution observations, offering the potential to identify and more accurately constrain substructure relative to CLEAN in low – moderate resolution datasets.

To further show that `frank` can identify substructure in lower resolution observations, **Fig. 14** is an animation in which the `frank` fit to the DSHARP AS 209 dataset evolves as a function of the data's longest baseline. We first truncate the data at 1 Mλ ⇔ ≈0.2″ prior to the visibilities being fit, then successively step the maximum baseline to 5 Mλ and fit at each step. `frank` identifies with increasing accuracy the presence and detailed morphology of substructure in the disc as the truncation baseline increases, and at each step, including when 1 Mλ is the longest baseline, the model is correctly identifying and partially resolving more structure than a CLEAN profile obtained with the same truncated dataset. This again demonstrates that `frank` can correctly identify structure in low – moderate resolution data.

`frank`'s resolving capability for low – moderate resolution observations also shows utility when applied to apparently smooth discs. For the FWHM 92 × 127 mas (≈ 15 AU at 140 pc) observations of the compact (≈0.3″) disc DR Tau (Long et al. 2019), the CLEAN image (even with uniform weighting) shows no clear signs of substructure, and the parametric visibility domain fit with a smooth profile
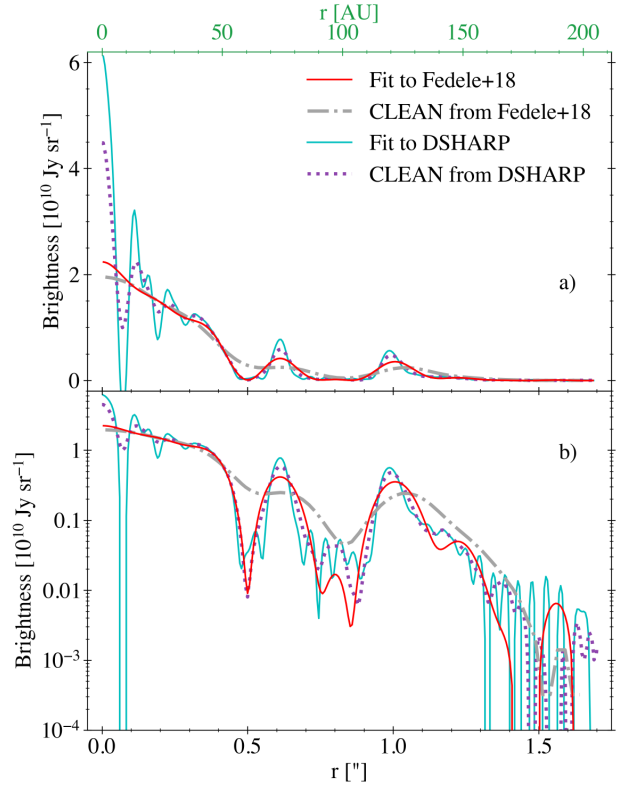
**Figure 13. Comparison of moderate and high resolution fits: AS 209**
a) The fitted `frank` brightness profile in Fig. 10(a), as well as a fit to lower resolution (longest baseline 2 Mλ) observations of the same disc (Fedele et al. 2018). Analogs for the CLEAN image-extracted profile. The `frank` fit to the lower resolution dataset is correctly identifying structure evident in both the `frank` and CLEAN fits to the higher resolution data.
b) As in (a) on a logarithmic y-scale.

in Long et al. (2019) shows significant residual error[8]. The `frank` fit using default hyperparameter values in **Fig. 15**(a) – (b) identifies substructure and finds a factor of ≈2 higher peak brightness than the CLEAN profile. This is because the `frank` fit matches the visibilities in Fig. 15(c) – (d) out to the longest baselines, 2.3 Mλ ⇔ 88 mas = 12 AU. Conversely convolution with the FWHM 92 × 127 mas synthesized beam causes the DHT of the CLEAN profile to depart the data beyond ≈0.8 Mλ. The DHT of the CLEAN profile does track at lower accuracy the same features in the visibility distribution that `frank` closely fits, and the `frank` fit when convolved with the synthesized beam in Fig. 15(a) – (b) matches the CLEAN profile. Together these suggest the `frank` fit is correctly identifying sub-beam structure.

`frank` thus infers the presence of substructure (two or more gaps) in addition to the centrally peaked component

---

[8] We used the self-calibrated measurement set in Long et al. (2019). Before extracting the visibilities using the `export_uvtable` function of the `uvplot` package (Tazzari 2017), we applied channel averaging (to obtain 1 channel per spectral window) and time averaging (30 sec) to all spectral windows in the original MS table.
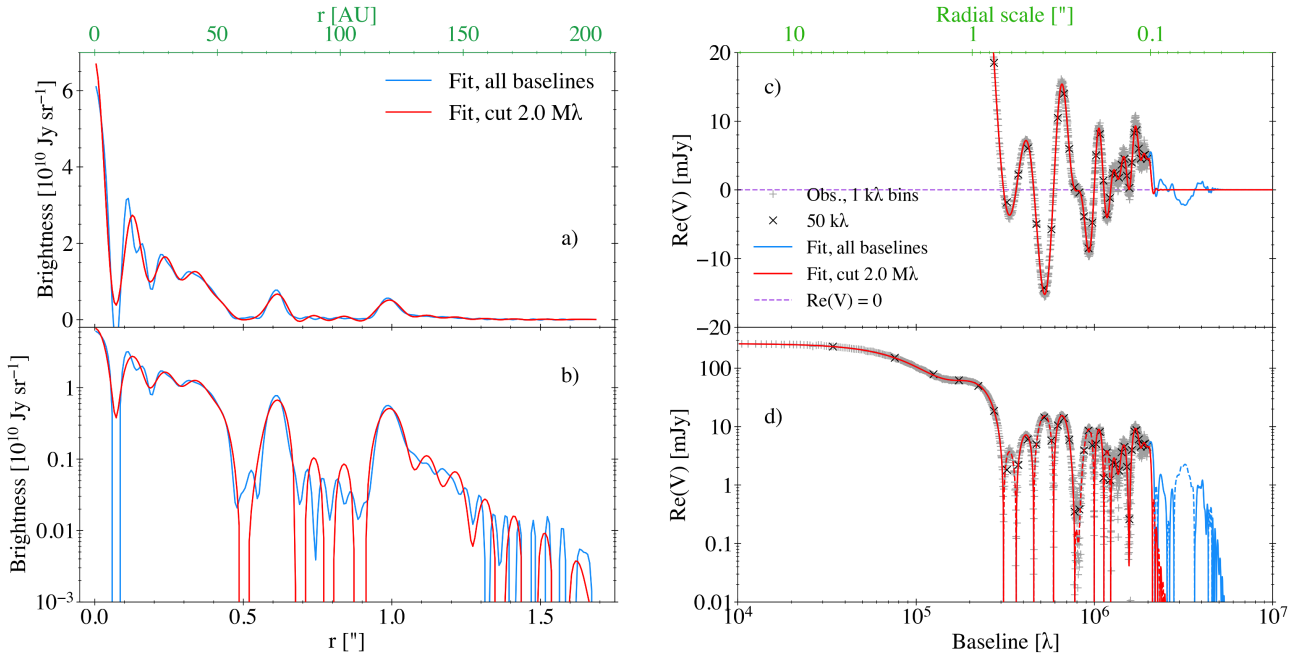
Figure 14. Animation of AS 209 fit evolution with baseline
a) The fitted `frank` brightness profile in Fig. 10(a) and the fit's evolution as the data are truncated at successively longer maximum baseline, beginning at 1 Mλ.
b) As in (a) on a logarithmic y-scale.
c) Zoom on the region around 0 of the observed DSHARP visibilities in 1 and 50 kλ bins. The `frank` fit to the full dataset and to the data at the evolving truncation baseline are shown.
d) As in (c) on a logarithmic y-scale. The data (and fits') negative regions are denoted by the fit lines' dashed sections.
The animated figure is shown in the online version; a snapshot at a truncation baseline of 2 Mλ is shown in the print version.

in this compact disc that appears featureless in the CLEAN image and profile. Note that `frank` is still almost certainly underresolving the disc's features, identifying and localizing the presence of substructure rather than accurately characterizing the features' morphologies. These are governed by the longest baselines in the dataset, so while it is clear from the current data that substructure is present, its form would likely change with the addition of higher resolution data (as is the case for AS 209). We therefore caution against over-interpreting the exact form of the fitted brightness profile.

Fig. 15(e) – (h) compare the CLEAN image with the image of the *unconvolved and deprojected* `frank` fit, the fit convolved with the beam, and the convolved residual image of the fit. The latter shows prominent asymmetric structure whose origin is not yet clear, but is not as best we can discern an artifact of an erroneous fit.

Varying the fit hyperparameters α and w_{smooth} within sensible bounds has a fairly weak effect on the `frank` brightness profile as shown in **Fig. 16**. While the stronger hyperparameter choice ($\alpha = 1.30, w_{\text{smooth}} = 10^{-1}$) does fit the visibilities to shorter baseline than the default hyperparameter values ($\alpha = 1.05, w_{\text{smooth}} = 10^{-4}$), the effect of using the stronger values is fairly benign: slightly less prominent disc features and a peak brightness $\approx$10% lower.

4   CONCLUSIONS

We have presented `frank`, an open source code that uses a fast Gaussian process to recover the axisymmetric structure of sources observed with radio interferometers by directly fitting the real component of the visibilities. This work focuses on protoplanetary disc observations, though the code can be extended to applications in other physical contexts where an azimuthally averaged brightness profile is useful.

`frank`'s main advantages for characterizing axisymmetric structure relative to the CLEAN technique are:

- resolution: `frank` can resolve angular scales smaller than the uniform-weighted CLEAN beam while retaining sensitivity typical of a natural-weighted CLEAN image;
- flexibility: `frank` yields a nonparametric reconstruction of a source's radial brightness profile, enabling it to fit the visibilities to high accuracy without additional input from the user;
- speed: `frank` fits are performed in <1 min for datasets of $\lesssim$ few × 10^6 visibilities;
- ease of use: `frank` is a Python package with a minimum number of dependencies that is easy to install and integrate into existing codebases.

`frank`'s resolution, flexibility, speed, and ease of use make it a high performance tool suitable for a wide range of applications on diverse interferometric datasets.

The major use case of `frank` in the protoplanetary disc context is fitting interferometric observations to find and
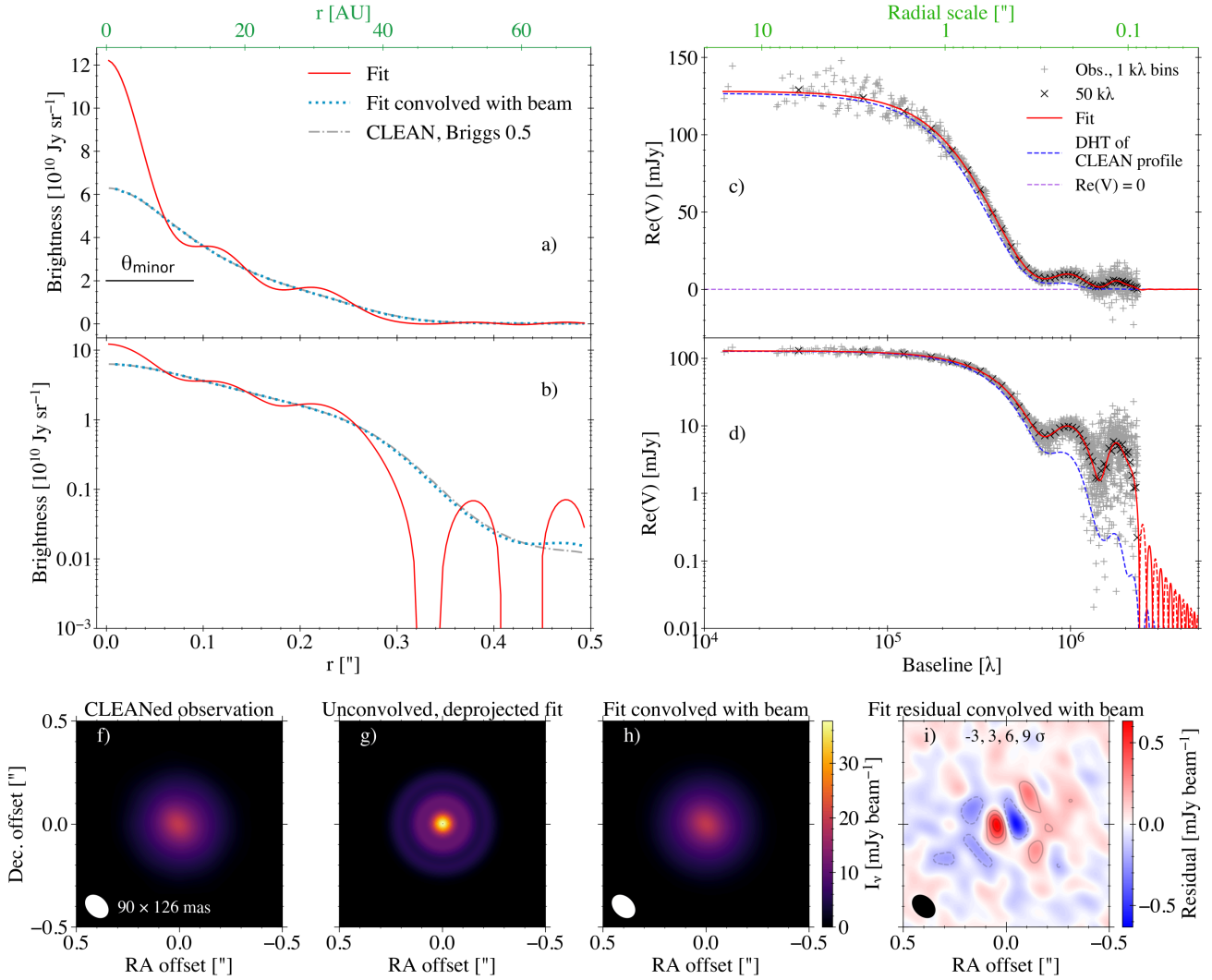
Figure 15. Fit to real, moderate resolution observations: DR Tau

a) Reconstructed brightness profile for the real, moderate resolution observations of DR Tau in Long et al. (2019). A CLEAN image-extracted profile and the beam's minor axis are also shown.

b) As in (a) on a logarithmic y-scale, emphasizing the fitted profile's outer edge at $\approx 0.3''$ (oscillations in the `frank` profile beyond this indicate the fit's noise floor).

c) Observed visibilities in 1 and 50 kλ bins. The `frank` fit and the discrete Hankel transform of the CLEAN profile in (a) are shown.

d) As in (c) on a logarithmic y-scale. The `frank` fit's negative regions are denoted by its dashed sections.

e) CLEANed observation.

f) Image of the *unconvolved and deprojected* `frank` reconstruction.

g) Image of the `frank` reconstruction reprojected and convolved with the observations' synthesized beam.

h) Residual map of the `frank` reconstruction (the fit's residual visibilities imaged with CLEAN), showing evidence of non-axisymmetric structure (contour $\sigma = 60$ μJy beam^{-1}, peak residual 12σ).

characterize disc substructures at higher resolution than the CLEAN algorithm. Recent high resolution observations of sources previously observed at lower angular resolution have begun to show the ubiquity of disc substructure, including in discs that appeared featureless at low resolution. This motivates the utility of a technique such as `frank` to resolve features and characterize discs at sub-beam scales. The model can resolve additional structure in high resolution datasets, such as the DSHARP observations of AS 209 (Andrews et al. 2018); as well as in lower resolution datasets that show featureless CLEAN images (even with uniform weighting), such

as the compact disc DR Tau (Long et al. 2018). Specific scientific applications of the code include more accurately discerning substructure widths and amplitudes, better resolving the inner disc structure including the peak brightness, and isolating disc asymmetries in a residual image.

`frank` uses all visibilities in a dataset to inform the reconstructed brightness profile. Its resolving power is thus sensitive to variations in the baseline dependence of the data's SNR. In contrast a CLEAN image-extracted profile is largely determined by the pure baseline distribution. Consequently for observations at any resolution, improving an
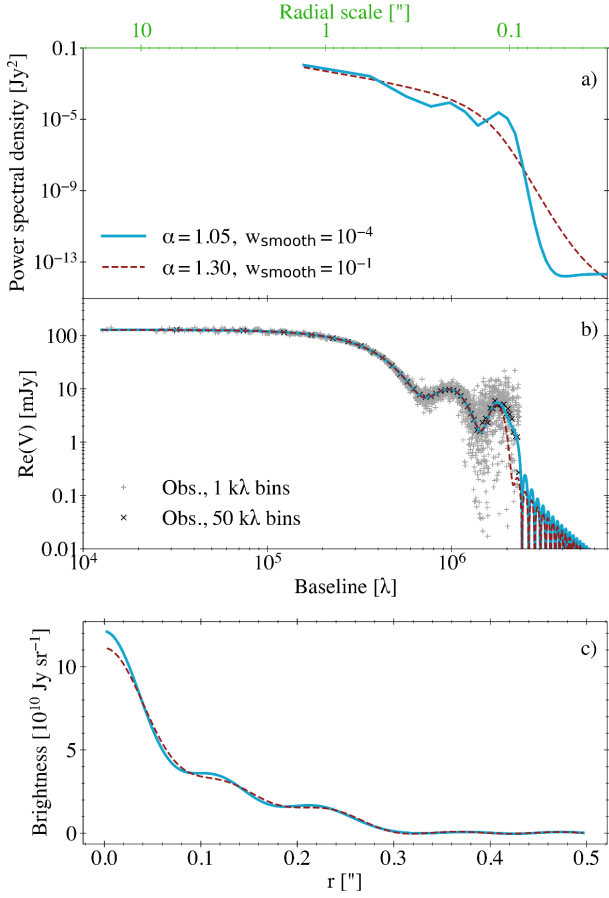
Figure 16. Fit sensitivity to hyperparameters: DR Tau
a) Maximum a posteriori power spectra of the `frank` fit under the default hyperparameter values ($\alpha = 1.05$, $w_{smooth} = 10^{-4}$) and stronger values ($\alpha = 1.30$, $w_{smooth} = 10^{-1}$). The structural differences between these spectra show weak effect on the visibility fits in (b) and fitted profiles in (c). The power spectra are the priors placed on the respective brightness profile reconstructions.
b) Observed visibilities in 1 and 50 kλ bins; the `frank` fit under the default hyperparameter values and under a stronger choice, showing fairly minor difference.
c) The fitted `frank` brightness profile in Fig. 15(a) – (b) under the default hyperparameter values, and a fit under a stronger choice.

observation's SNR via the on-source integration time can markedly enhance `frank`'s resolving capability to constrain sub-beam features, whereas it makes little difference for CLEAN.

`frank` is actively developed at https://github.com/discsim/frank, where users are welcome to contribute and to report issues.

REFERENCES

ALMA Partnership T., et al., 2015, The Astrophysical Journal Letters, 808, L3
Ables J. G., 1974, A&AS, 15, 383
Abrantes F., Lopes C., Rodrigues T., Gil I., Witt L., Grimalt J., Harris I., 2009, Geochemistry, Geophysics, Geosystems, 10, Q09U07
Akiyama K., Matthews L. D., 2019, Exploring Regularized Maximum Likelihood Reconstruction for Stellar Imaging with the ngVLA (arXiv:1910.00013)
Alexander R. D., Clarke C. J., Pringle J. E., 2006, Monthly Notices of the Royal Astronomical Society, 369, 229âĂŞ239
Andrews S. M., et al., 2016, The Astrophysical Journal, 820, L40
Andrews S. M., et al., 2018, ApJ, 869, L41
Arras P., Knollmüller J., Junklewitz H., Enßlin T. A., 2018, arXiv e-prints, p. arXiv:1803.02174
Arras P., Frank P., Leike R., Westermann R., Enßlin T. A., 2019, A&A, 627, A134
Baddour N., Chouinard U., 2015, Journal of the Optical Society of America A, 32, 611
Bracewell R. N., 2000, The Fourier transform and its applications
Chael A. A., Johnson M. D., Narayan R., Doeleman S. S., Wardle J. F. C., Bouman K. L., 2016, The Astrophysical Journal, 829, 11
Cieza L. A., et al., 2019, MNRAS, 482, 698
Clark B. G., 1980, A&A, 89, 377
Clarke C. J., Gendrin A., Sotomayor M., 2001, Monthly Notices of the Royal Astronomical Society, 328, 485
Clarke C. J., et al., 2018, ApJ, 866, L6
Cornwell T. J., Evans K. F., 1985, A&A, 143, 77
Cornwell T., Braun R., Briggs D. S., 1999, in Taylor G. B., Carilli C. L., Perley R. A., eds, Astronomical Society of the Pacific Conference Series Vol. 180, Synthesis Imaging in Radio Astronomy II. p. 151
Dullemond C. P., et al., 2018, ApJ, 869, L46
Enßlin T. A., Frommert M., 2011, Phys. Rev. D, 83, 105014
Event Horizon Telescope Collaboration et al., 2019, ApJ, 875, L4
Fedele D., et al., 2017, A&A, 600, A72
Fedele D., et al., 2018, A&A, 610, A24
Flock M., Ruge J. P., Dzyurkevich N., Henning T., Klahr H., Wolf S., 2015, A&A, 574, A68
Flock M., Nelson R. P., Turner N. J., Bertrang G. H.-M., Carrasco-González C., Henning T., Lyra W., Teague R., 2017, ApJ, 850, 131
Greiner M., Vacca V., Junklewitz H., Enßlin T. A., 2016, arXiv e-prints, p. arXiv:1605.04317
Gull S. F., Daniell G. J., 1978, Nature, 272, 686
Guzmán V. V., et al., 2018, ApJ, 869, L48
Haffert S. Y., Bohn A. J., de Boer J., Snellen I. A. G., Brinchmann J., Girard J. H., Keller C. U., Bacon R., 2019, Nature Astronomy, p. 329

Hallam P. D., Paardekooper S. J., 2020, MNRAS, 491, 5759

Högbom J. A., 1974, A&AS, 15, 417

Honma M., Akiyama K., Uemura M., Ikeda S., 2014, Publications of the Astronomical Society of Japan, 66

Huang J., et al., 2018, ApJ, 869, L42

Hunter J. D., 2007, Computing in Science & Engineering, 9, 90

Isella A., Benisty M., Teague R., Bae J., Keppler M., Facchini S., PÃl'rez L., 2019, The Astrophysical Journal, 879, L25

Johnson M. D., et al., 2017, The Astrophysical Journal, 850, 172

Junklewitz H., Bell M. R., Enßlin T., 2015, A&A, 581, A59

Junklewitz H., Bell M. R., Selig M., Enßlin T. A., 2016, A&A, 586, A76

Keppler M., et al., 2019, Astronomy & Astrophysics, 625, A118

Kley W., Nelson R. P., 2012, ARA&A, 50, 211

Kluyver T., et al., 2016, in Loizides F., Schmidt B., eds, Positioning and Power in Academic Publishing: Players, Agents and Agendas. pp 87 – 90

Kudo T., Hashimoto J., Muto T., Liu H. B., Dong R., Hasegawa Y., Tsukagoshi T., Konishi M., 2018, ApJ, 868, L5

Lin D. N. C., Papaloizou J., 1979, MNRAS, 188, 191

Long F., et al., 2018, ApJ, 869, 17

Long F., et al., 2019, ApJ, 882, 49

Lu R.-S., Broderick A. E., Baron F., Monnier J. D., Fish V. L., Doeleman S. S., Pankratius V., 2014, The Astrophysical Journal, 788, 120

McMullin J. P., Waters B., Schiebel D., Young W., Golap K., 2007, CASA Architecture and Applications. p. 127

Meru F., Juhász A., Ilee J. D., Clarke C. J., Rosotti G. P., Booth R. A., 2017, ApJ, 839, L24

Narayan R., Nityananda R., 1986, ARA&A, 24, 127

Oppermann N., Selig M., Bell M. R., Enßlin T. A., 2013, Phys. Rev. E, 87, 032136

Pérez L. M., et al., 2016, Science, 353, 1519

Pérez S., Casassus S., Baruteau C., Dong R., Hales A., Cieza L., 2019, arXiv e-prints,

Pinilla P., Birnstiel T., Ricci L., Dullemond C. P., Uribe A. L., Testi L., Natta A., 2012, A&A, 538, A114

Pinte C., et al., 2019, Nature Astronomy, p. 419

Rasmussen C. E., Williams C. K. I., 2006, Gaussian Processes for Machine Learning

Remijan A., et al., 2019, ALMA Doc. 7.3, ver. 1.0

Rosotti G. P., Juhasz A., Booth R. A., Clarke C. J., 2016, MNRAS, 459, 2790

Sutter P. M., et al., 2013, Monthly Notices of the Royal Astronomical Society, 438, 768

Tazzari M., 2017, mtazzari/uvplot: v0.1.1, doi:10.5281/zenodo.1003113, https://doi.org/10.5281/zenodo.1003113

Tazzari M., Beaujean F., Testi L., 2018, MNRAS, 476, 4527

Thompson A. R., Moran J. M., Swenson Jr. G. W., 2017, Interferometry and Synthesis in Radio Astronomy, 3rd Edition, doi:10.1007/978-3-319-44431-4.

Virtanen P., et al., 2019, SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python (arXiv:1907.10121)

Walt S. v. d., Colbert S. C., Varoquaux G., 2011, Computing in Science & Engineering, 13, 22

Zhang K., Blake G. A., Bergin E. A., 2015, ApJ, 806, L7

Zhang K., Bergin E. A., Blake G. A., Cleeves L. I., Hogerheijde M., Salinas V., Schwarz K. R., 2016, ApJ, 818, L16

APPENDIX A: SPECTRAL SMOOTHNESS HYPERPRIOR

The spectral smoothness hyperprior is designed to constrain the power spectrum to be close to a power law in the absence of data. We implement this hyperprior using a numerical estimate for the integral of the second logarithmic derivative of the power spectrum,

$$-\frac{1}{2\sigma_s^2}\int d\log(q)\left(\frac{\partial^2\log(p)}{\partial\log(q)^2}\right)^2 \approx -\frac{1}{2\sigma_s^2}\sum_{n=2}^{N-1}\delta_{c,n}(\Delta\log p_n)^2, \tag{A1}$$

as in Oppermann et al. (2013). Here $\delta_{c,n} = (\log q_{n+1} - \log q_{n-1})/2$ and

$$\Delta\log p_n = \frac{1}{\delta_{c,n}}\left(\frac{\log p_{n+1}-\log p_n}{\delta_{e,n}} - \frac{\log p_n - \log p_{n-1}}{\delta_{e,n-1}}\right) \tag{A2}$$

is an approximation to the second logarithmic derivative of p, where $\delta_{e,n} = \log q_{n+1} - \log q_n$. Equation A1 can be simplified as

$$-\frac{1}{2\sigma_s^2}\sum_{n=2}^{N-1}\delta_{c,n}(\Delta\log p_n)^2 = -\frac{1}{2\sigma_s^2}\log p^T\, \mathbf{T}\, \log p, \tag{A3}$$

where the components of $\mathbf{T}$ are

$$T_{ij} = \sum_k \gamma_{ki}\delta_{c,k}\gamma_{kj}, \tag{A4}$$

with

$$\gamma_{ij} = \frac{1}{\delta_{c,i}}\begin{cases} -\left(\frac{1}{\delta_{e,i}}+\frac{1}{\delta_{e,i-1}}\right) & \text{if } j = i \\ +\frac{1}{\delta_{c,i\pm1}} & \text{if } j = i\pm1 \\ 0 & \text{otherwise.} \end{cases} \tag{A5}$$

The matrix $\mathbf{T}$ is constant and pentadiagonal, depending only on q_k. For large k, $\mathbf{T}$ has the asymptotic form

$$T_{ij} = \frac{1}{(q_iR_{\max})^3}\times\begin{cases} +48 & \text{if } j = i \\ -32 & \text{if } j = i\pm1 \\ +8 & \text{if } j = i\pm2 \\ 0 & \text{otherwise.} \end{cases} \tag{A6}$$

APPENDIX B: SIGNAL-TO-NOISE THRESHOLD

Here we show that it is the visibilities' SNR that determines whether a `frank` fit ignores the data at a given baseline. This is ultimately determined by whether the maximum a posteriori value for the power spectrum parameters tends to: Equation 43, in which case the power spectrum is determined by the visibility amplitude; or Equation 44, in which case the power spectrum's low amplitude suppresses the power on a given scale. The argument presented here is similar to that in Enßlin & Frommert (2011), who derive this for a general class of methods like `frank`.

To make the derivation tractable, we make some simplifications: we assume $p_0 = 0$, neglect smoothing, and study the simplifying case where the visibilities are observed exactly at the spatial frequency collocation points q_k. While this is somewhat unrealistic, the insights derived are useful more generally. Under our last assumption, the matrix $\mathbf{H}(q)$ in Equation 24 is just $\mathbf{Y}_f$. Now,

$$\begin{aligned}\mu &= \left(\mathbf{Y}_f^T\mathbf{N}^{-1}\mathbf{Y}_f + \mathbf{Y}_f^T\,\text{diag}\,(1/p)\mathbf{Y}_f\right)^{-1}\mathbf{Y}_f^T\mathbf{N}^{-1}V \\ &= \mathbf{Y}_b\left(\mathbf{N}^{-1}+\text{diag}\,(1/p)\right)^{-1}\mathbf{N}^{-1}V.\end{aligned} \tag{B1}$$

This means that

$$(\mathbf{Y}_{\mathrm f}\boldsymbol{\mu})_k = \frac{V_k}{1 + \sigma_k^2/p_k} \tag{B2}$$

and

$$(\mathbf{Y}_{\mathrm f}^T\mathbf{D}\mathbf{Y}_{\mathrm f})_{kk} = \frac{\sigma_k^2}{1 + \sigma_k^2/p_k}. \tag{B3}$$

Using these in Equation 42 with $p_0 = 0$, $\sigma_{\mathrm s} = \infty$, and $\boldsymbol{p}^{\mathrm{new}} = \boldsymbol{p}$ (convergence to the maximum likelihood) and rearranging produces the cubic equation in p_k,

$$p_k \left\{ (p_k + \sigma_k^2)^2 [1 + 2(\alpha - 1)] - (p_k + \sigma_k^2)(V_k^2 + \sigma^2) - V_k^2\sigma^2 \right\} = 0, \tag{B4}$$

which always has the solution $p_k = 0$. Solutions for $p_k > 0$ can be found by completing the square of the term in brackets;

$$p_k = \frac{1}{1 + 2(\alpha - 1)} \times$$
$$\frac{1}{2} \left\{ \left[V_k^2 - \sigma^2(1 + 4(\alpha - 1)) \right] + \right.$$
$$\left. \sqrt{(V_k^2 - \sigma_k^2)^2 - 8V_k^2\sigma_k^2(\alpha - 1)} \right\}. \tag{B5}$$

For $V_k \gg \sigma_k$, this yields an equivalent expression to that given in Equation 43. However, the term inside the square root in Equation B5 is only positive if

$$V_k^2 > \sigma^2 \left\{ 1 + 4(\alpha - 1) + 2\sqrt{2(\alpha - 1)[1 + 2(\alpha - 1)]} \right\}, \tag{B6}$$

otherwise $p_k = 0$ is the only solution (note that in the special case $\alpha = 1$, the term in brackets in Equation B5 is always positive; however the above expression still correctly denotes the region for which solutions with $p_k > 0$ exist). The implication of this is that for an SNR below a given threshold, $V_k^2 \lesssim \sigma_k^2$, the inverse Γ hyperprior will drive the power spectrum toward zero. In practice, including a nonzero p_0 means that $p_k \rightarrow p_0/(\alpha - 1)$, as given in Equation 44.

Equation B5 also shows why we use $\alpha > 1$. At long baselines – where the visibilities are noise-dominated – we will have $V_k^2 \sim \sigma_k^2$, and there will be many statistical fluctuations causing V_k to be slightly greater than σ_k. Since for $\alpha = 1$ the threshold is exactly at $V_k = \sigma_k$, the power spectrum will contain some fraction of this power, and the model will fit the noise-dominated visibilities. Increasing α to 1.05, 1.10, 1.30, or 1.90 increases the noise threshold to 1.36, 1.54, 2.04, or 3.01 σ_k respectively. Thus using $\alpha > 1$ markedly reduces the chance that the model will attempt to fit noise-dominated data.

A similar argument can be made in the case where the visibilities are not observed directly at the spatial frequency collocation points, so long as the (u, v) plane is sampled well enough that $\mathbf{M}$ can be approximated as a diagonal matrix in the visibility domain. In this case $V_k = \sigma_k^2 (\mathbf{Y}_{\mathrm b}^T \boldsymbol{j})_k$, where $\sigma_k^2 \approx (\mathbf{Y}_{\mathrm f}^T\mathbf{M}\mathbf{Y}_{\mathrm f})_{kk}^{-1}$ is the effective variance at that scale.

APPENDIX C: NONNEGATIVE FITS

In some circumstances it may be beneficial to have brightness profiles that are nonnegative, i.e., $\boldsymbol{I}_\nu \geq 0$, as is produced
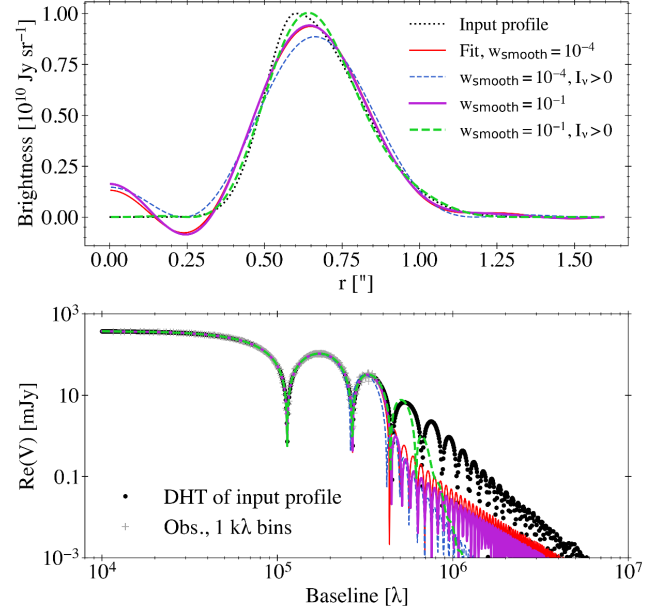
Figure C1. Comparison of standard and nonnegative fits
a) Input and reconstructed brightness profiles for a mock Gaussian ring (two joined sigmoids) observed with the ALMA C43-3 configuration (synthesized beam FWHM 0.59×0.70″, Briggs=0.5; see Table 2 and Sec. 3.1.1), shown under two values of the w_{smooth} hyperparameter, and with brightness profile positivity either unenforced or enforced. All fits use $\alpha = 1.05$.
b) Observed visibilities in 1 kλ bins, the DHT of the noiseless input profile, and the `frank` fits corresponding to the profiles in (a).

by maximum entropy methods or a log-normal model for the brightness (Junklewitz et al. 2016). In `frank` we can generate nonnegative solutions by finding the most probable brightness reconstruction $\boldsymbol{I}_\nu$ for a given set of power spectrum parameters and the constraint $\boldsymbol{I}_\nu \geq 0$ (i.e., the maximum of $P(\boldsymbol{I}_\nu|\boldsymbol{p})$ subject to $\boldsymbol{I}_\nu \geq 0$) using the nonnegative least squares solver in the SCIPY package (`scipy.optimize.nnls`).

In Fig. C1 we show the impact of the $\boldsymbol{I}_\nu \geq 0$ constraint on the reconstructed brightness profile for the mock Gaussian ring under the C43-3 mock observation. Comparing the posterior mean brightness profile and the most probable nonnegative solution under two values of w_{smooth}, Fig. C1(a) shows the nonnegative solution is more accurate. Correspondingly the constraint $\boldsymbol{I}_\nu \geq 0$ yields a visibility fit in Fig. C1(b) that initially falls slower than the standard `frank` fit beyond the data's longest baseline (0.4 Mλ), predicting the visibility amplitude at unobserved baselines more accurately than the standard fit between 0.4 – 0.8 Mλ.

The improvement gained by enforcing $\boldsymbol{I}_\nu \geq 0$ does however depend on the hyperparameters, with the nonnegative fit using $w_{\mathrm{smooth}} = 10^{-4}$ fitting the visibilities less accurately than the analogous standard `frank` fit. This hyperparameter dependency is a result of the need to introduce additional power on small scales when enforcing a nonnegative brightness. Large α and small w_{smooth} strongly damp power at long baseline (on small spatial scales), introducing tension between the prior and the $\boldsymbol{I}_\nu \geq 0$ constraint. This can cause the nonnegative reconstruction to be a poor fit to the

data on shorter baselines. Conversely with smaller α and larger w_{smooth}, the maximum a posteriori power spectra fall off more slowly with increasing q, and the resulting non-negative brightness reconstruction has more power on small spatial scales.

Because this process involves extrapolating the visibilities beyond their maximum baseline (or at least into a region where they are noise-dominated), we therefore urge caution when interpreting any additional structures introduced by enforcing the profile be nonnegative.

APPENDIX D: MOCK DISC FUNCTIONAL FORMS

Table D1 gives the functional forms of the brightness profiles for the mock discs shown in Fig. 7 and listed in Table 2.

This paper has been typeset from a TEX/LATEX file prepared by the author.

Table D1. Functional forms of the brightness profiles for the archetypal mock discs listed in Table 2 and shown in Fig. 7.

Disc	Brightness profile functional form	Amplitude variables $[10^{10} \text{ Jy sr}^{-1}]$	Location variables ["]
Gaussian centered at 0	$I_\nu(r) = ae^{\frac{-r^2}{2c^2}}$	$a = 1.0$	$c = 0.2$
Gaussian ring	$I(r) = \begin{cases} ae^{\frac{-(r-b)^2}{2c_1^2}}, & r \leq b \\ ae^{\frac{-(r-b)^2}{2c_2^2}}, & r > b \end{cases}$	$a = 1.0$	$b = 0.6$ $c_1 = 0.1$ $c_2 = 0.2$
Multi-Gaussian	$I(r) = a_1 e^{\frac{-(r-b_1)^2}{2c_1^2}} - a_2 e^{\frac{-(r-b_2)^2}{2c_2^2}} - a_3 e^{\frac{-(r-b_3)^2}{2c_3^2}} + \begin{cases} a_4 e^{\frac{-(r-b_4)^2}{2c_4^2}}, & r \leq b_4 \\ a_4 e^{\frac{-(r-b_4)^2}{2c_5^2}}, & r > b_4 \end{cases}$	$a_1 = 0.6$ $a_2 = 0.4$ $a_3 = 0.3$ $a_4 = 1.0$	$b_1 = 0.70$ $b_2 = 0.60$ $b_3 = 0.85$ $b_4 = 0.15$ $c_1 = 0.20$ $c_2 = 0.05$ $c_3 = 0.05$ $c_4 = 0.05$ $c_5 = 0.10$
Sharp-edged	$I(r) = e^{-2r^2} \cdot \left(a_1 + \begin{cases} a_2, & b_1 \leq r \leq b_2 \\ a_3, & b_3 \leq r \leq b_4 \\ a_4, & b_5 \leq r \leq b_6 \\ a_5, & b_7 \leq r \leq b_8 \\ 0, & \text{else} \end{cases} \right)$	$a_1 = 1.00$ $a_2 = -0.90$ $a_3 = -0.20$ $a_4 = -0.25$ $a_5 = 0.50$	$b_1 = 0.1$ $b_2 = 0.2$ $b_3 = 0.4$ $b_4 = 0.7$ $b_5 = 0.6$ $b_6 = 0.7$ $b_7 = 0.7$ $b_8 = 0.8$