

Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging

Tobías I. Liaudat^{1,2,3*}, Matthijs Mars^{1,2}, Matthew A. Price², Marcelo Pereyra^{4,5}, Marta M. Betcke¹ and Jason D. McEwen^{2,6*}

¹Department of Computer Science, University College London (UCL), London WC1E 6BT, UK

²Mullard Space Science Laboratory (MSSL), University College London (UCL), Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

³IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

⁴School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK

⁵Maxwell Institute for Mathematical Sciences, Bayes Centre, 47 Potterrow, Edinburgh EH8 9BT, UK

⁶Alan Turing Institute, Euston Road, London NW1 2DB, UK

Accepted 2024 July 31. Received 2024 July 24; in original form 2023 December 4

ABSTRACT

Next-generation radio interferometers like the Square Kilometer Array have the potential to unlock scientific discoveries thanks to their unprecedented angular resolution and sensitivity. One key to unlocking their potential resides in handling the deluge and complexity of incoming data. This challenge requires building radio interferometric (RI) imaging methods that can cope with the massive data sizes and provide high-quality image reconstructions with uncertainty quantification (UQ). This work proposes a method coined QUANTIFAI to address UQ in RI imaging with data-driven (learned) priors for high-dimensional settings. Our model, rooted in the Bayesian framework, uses a physically motivated model for the likelihood. The model exploits a data-driven convex prior potential, which can encode complex information learned implicitly from simulations and guarantee the log-concavity of the posterior. We leverage probability concentration phenomena of high-dimensional log-concave posteriors to obtain information about the posterior, avoiding MCMC sampling techniques. We rely on convex optimization methods to compute the MAP estimation, which is known to be faster and better scale with dimension than MCMC strategies. QUANTIFAI allows us to compute local credible intervals and perform hypothesis testing of structure on the reconstructed image. We propose a novel fast method to compute pixel-wise uncertainties at different scales, which uses three and six orders of magnitude less likelihood evaluations than other UQ methods like length of the credible intervals and Monte Carlo posterior sampling, respectively. We demonstrate our method by reconstructing RI images in a simulated setting and carrying out fast and scalable UQ, which we validate with MCMC sampling. Our method shows an improved image quality and more meaningful uncertainties than the benchmark method based on a sparsity-promoting prior.

Key words: Radio Interferometric Imaging – Machine Learning – Algorithms – Data Methods.

1 INTRODUCTION

Radio astronomy plays a crucial role in expanding our understanding of the Universe, offering a unique perspective on astrophysical and cosmological phenomena. Among the transformative tools in an astronomer's toolkit, radio interferometric (RI) imaging stands out as an indispensable technique. Aperture synthesis and radio interferometry (Thompson, Moran & Swenson 2017) allow us to achieve high angular resolutions providing immense power to resolve objects. Furthermore, radio frequency signals are only weakly attenuated by our atmosphere, allowing for observations at the Earth's surface. The unparalleled angular resolution, high sensitivity, and the different phenomena emitting in the radio wavelength regime make RI an ideal candidate to better help us understand our Universe.

The advent of the Square Kilometre Array (SKA; Dewdney et al. 2009) heralds a new era in radio astronomy (Braun et al. 2015) spanning the study from the epoch of reionization and fast radio bursts to galaxy evolution and dark energy. SKA's vast collecting area and sensitivity promise to provide a leap forward in our observational capabilities, opening doors to discoveries. However, this transformative potential comes with the formidable computational challenge of processing and making sense of the unprecedented volume of SKA-generated data. Developing and implementing algorithms that can efficiently handle SKA's data deluge is a challenge. In addition, achieving the high reconstruction performance required to unlock SKA's full potential is a significant obstacle in the SKA's data processing requirements.

The aperture synthesis techniques in RI probe the sky by acquiring specific Fourier measurements, which results in incomplete coverage of the Fourier domain of the sky's image of interest. The incomplete Fourier coverage makes the problem of estimating the underlying sky image, which we know as RI imaging, an ill-posed inverse

* E-mail: tobiasliaudat@gmail.com (TIL); jason.mcewen@ucl.ac.uk (JDM)

problem, which is further complexified by the observational noise. Having a way to quantify the uncertainty in the image reconstructions becomes essential given the uncertainties involved in the RI imaging problem. To make scientifically sound inferences and informed decisions, we need the ability to quantify these uncertainties rigorously. This motivates the development of uncertainty quantification (UQ) methods tailored to the complexities of radio interferometric data, where scalability, i.e. the computational complexity with respect to the amount of data processed, and performance play a central role. We need to ensure that our reconstructions are not only insightful but also trustworthy.

In a nutshell, we want to develop RI imaging methods that can deliver precision with uncertainty quantification and that are highly scalable. Most existing methods only tackle some of these three requirements. The widely used CLEAN algorithm (Högbom 1974) built its success on scalability and fast inference. CLEAN and its extensions (Cornwell 2008; Offringa et al. 2014; Offringa & Smirnov 2017) have been continuously used in many RI imaging pipelines since its inception. Despite offering limited imaging quality and reconstruction artefacts compared to other approaches, CLEAN stands out due to its scalability. More recent approaches leverage compressed sensing theory, relying on sparse priors (often in wavelet representations) and convex optimization techniques (Wiaux et al. 2009; McEwen & Wiaux 2011; Carrillo, McEwen & Wiaux 2012, 2014; Dabbech et al. 2015, 2018; Pratley et al. 2018). These methods have been shown to improve the reconstruction quality at the expense of increased computational complexity. Considerable work has been directed to parallelization and acceleration efforts for sparsity-based methods (Onose et al. 2016; Pratley et al. 2019a; Pratley, Johnston-Hollitt & McEwen 2019b; Thouvenin et al. 2022a, b).

The deep learning revolution has introduced a powerful way to encode complex image priors in neural networks, which may be used to solve complex high-dimensional inverse problems. This data-driven or learned paradigm has gained much traction across imaging landscape problems, including RI imaging (Allam 2016; Terris et al. 2022; Aghabiglou et al. 2023; Mars, Betcke & McEwen 2023). Learned methods can improve the reconstruction quality with respect to handcrafted priors, such as sparsity-based wavelet priors, as well as provide acceleration (Terris et al. 2022; Mars et al. 2023, 2024; Aghabiglou et al. 2024) to convex optimization-based methods.

Unfortunately, none of the RI imaging methods mentioned above, learned, sparsity-based, or CLEAN-based, provide UQ tools for a given model. Cai, Pereyra & McEwen (2018a, b) proposed methods for Bayesian UQ on RI imaging problems. Cai et al. (2018a) leverages proximal Markov chain Monte Carlo (MCMC) methods (Pereyra 2016) to provide support for sparsity-promoting priors. The proposed method allows them to reconstruct the image and provide UQ by sampling the posterior probability distribution. The drawback of the method is the high computational cost suffered by all MCMC sampling techniques. The companion paper, Cai et al. (2018b), overcomes the need for posterior sampling with maximum-a-posteriori (MAP) based UQ (Pereyra 2017) relying on convex optimization techniques. The second method (Cai et al. 2018b) provides a significant speed-up with respect to the sampling-based method (Cai et al. 2018a), but its reconstruction quality is limited to sparsity-promoting priors.

Other RI imaging methods have addressed UQ for their reconstructions. Dia et al. (2023) proposed to use score-based generative models as priors (Ho, Jain & Abbeel 2020; Song et al. 2020), which by employing the convolved likelihood approximation (Adam et al. 2022; Remy et al. 2023) are able to sample from the posterior. How-

ever, the method is computationally costly, and the sampling relies on MCMC methods. The Bayesian RI imaging method `comrade` (Tiede 2022) was developed for very-long-baseline interferometry (VLBI) aiming to image black holes and active galactic nuclei. The `comrade` method relies on MCMC sampling methods like nested sampling (Ashton et al. 2022) and Hamiltonian Monte Carlo (Xu et al. 2020) to sample from the posterior. Posterior sampling based on MCMC methods can be enough for VLBI but cannot yet cope with the dimensions of SKA-like images.

An alternative Bayesian RI imaging algorithm is the `resolve` method and its upgrades (Arras et al. 2018, 2021, 2022; Knollmüller, Arras & Enßlin 2023; Roth et al. 2023), introduced initially in Junklewitz et al. (2016), which can produce uncertainty maps from approximate posterior samples. The model is based on Gaussian random fields with different analytical priors used to promote certain aspects of the reconstruction, like positivity and spatial and temporal correlations. One novelty of the method is to approximate the true posterior probability distribution with variational inference and, therefore, be able to sample from the approximated distribution without resorting to MCMC methods that do not scale well in very high dimensions. The approximated posterior follows a multivariate Gaussian distribution, where their parameters are learnt by trying to maximize the information overlap between the true posterior and the model by minimizing a Kullback-Leibler (KL-) divergence. Given that the full covariance matrix scales with the squared of the number of pixels, the authors exploit an approximation in the vicinity of the mean estimate. The method is based on the metric Gaussian variational inference (MGVI) approach from Knollmüller & Enßlin (2019). `resolve` has been used for reconstructions with VLBI observations of M87* (Arras et al. 2022) and Sagittarius A* (Knollmüller et al. 2023) from the Event Horizon Telescope (EHT) (Event Horizon Telescope Collaboration 2019a, b), as well as observations of Cygnus A from the very large array (VLA) (Arras et al. 2018, 2021; Roth et al. 2023).

In this article, we delve into the forefront of RI imaging and propose a method coined QUANTIFAI, based on learned convex priors, capable of delivering high-quality reconstructions with uncertainty quantification and being highly scalable. The method relies on the mathematically principled Bayesian framework to provide an understanding of the uncertainties through the posterior distribution. By restricting our model to log-concave posteriors, we can exploit recent MAP-based UQ techniques (Pereyra 2017), providing scalable optimization-based UQ. We build upon recent advances in neural-network-based convex regularizers (Goujon et al. 2023b), allowing us to improve the reconstruction quality and obtain more meaningful uncertainties with respect to Cai et al. (2018b). On top of the hypothesis tests of structure on the reconstructed image, we propose a novel fast method to estimate pixel-wise uncertainties as a function of scale.

The remainder of this article is organized as follows. In Section 2, we start by reviewing the RI imaging and techniques for the resulting inverse problem. Section 3 describes QUANTIFAI, the proposed method, and the RI image reconstruction algorithm. In Section 4, we introduce the core of our scalable UQ and the different UQ techniques it allows. The experimental results, including the performance of QUANTIFAI reconstruction and its UQ techniques, are presented in Section 5. Section 6 presents experimental results using more realistic observation based on simulated ungridded visibility patterns from the MeerKAT radio telescope. In Section 7 we provide a discussion on some limitations and possible extensions of the proposed methodology. We provide concluding remarks and present some future perspectives in Section 8.

2 RADIO INTERFEROMETRIC IMAGING

In this section, we start by reviewing the RI imaging inverse problem and discuss approaches to tackle it, including sparsity-based regularization, the CLEAN method, and learned approaches. We then introduce the Bayesian framework elements needed such as MAP estimation and proximal MCMC sampling algorithms that will be later used as validation.

2.1 Radio interferometry

The interferometric measurement equation for a radio telescope (Thompson et al. 2017) in the monochromatic setting relates our observations represented by the visibility function \mathcal{Y} to the sky brightness \mathcal{X} , which we want to reconstruct,

$$\mathcal{Y}(u, v, w) = \iint \frac{\mathcal{X}(l, m) \mathcal{A}(l, m)}{\sqrt{1-l^2-m^2}} \times \exp \left[-2\pi i w \left(\sqrt{1-l^2-m^2} - 1 \right) \right] \exp[-2\pi i (lu + mv)] dl dm, \quad (1)$$

where $\mathbf{u} = (u, v, w)$ are the interferometer baseline coordinates with units depending on the observation wavelength, $\mathbf{l} = (l, m, n)$ are cosine sky coordinates restricted to the unit sphere, and \mathcal{A} includes direction-dependent effects (DDEs) like the primary beam of the dishes. The previous general model allows us to consider different DDEs through \mathcal{A} and non-coplanar effects through the exponential term in w . These effects become considerable when considering wide fields of view and long baselines. There exists a rich body of literature incorporating such effects, e.g. Smirnov (2011a, b, c, d), Thompson et al. (2017), and there are scalable algorithms that take them into account, e.g. Pratley et al. (2019b).

In this article, for the sake of simplicity but without loss of generality, we assume the coplanar setting, where the antennas are located in the same w plane. We also assume that we observe a small field of view such that $1-l^2-m^2 \approx 1$. Consequently, we have $\exp[-2\pi i w (\sqrt{1-l^2-m^2} - 1)] \approx 1$, and equation (1) reduces to

$$\mathcal{Y}(u, v) \approx \iint \mathcal{X}(l, m) \mathcal{A}(l, m) \exp[-2\pi i (lu + mv)] dl dm. \quad (2)$$

From the previous equation, we can notice the remarkable result of $\mathcal{Y}(u, v) = \mathcal{F}(\mathcal{A}\mathcal{X})(u, v)$ where \mathcal{F} is the 2D Fourier transform.

To further simplify the problem, we will avoid using the continuous \mathcal{Y} and \mathcal{X} and work with their discrete counterparts, \mathbf{x} and \mathbf{y} , respectively. The observational model we study for our RI imaging problem writes

$$\mathbf{y} = \Phi \mathbf{x} + \mathbf{n}, \quad (3)$$

where $\mathbf{y} \in \mathbb{C}^M$ are the M observed complex visibilities, $\mathbf{x} \in \mathbb{R}^N$ is the discrete sky brightness sampled on a N point grid, and $\Phi \in \mathbb{C}^{M \times N}$ is the linear measurement operator that models the acquisition process. Without loss of generality, the observational and instrumental noise $\mathbf{n} \in \mathbb{C}^M$ is assumed to be independent and identically distributed (iid) white Gaussian noise with zero mean and standard deviation σ . If the noise is not white, we can incorporate a noise whitening matrix in the Φ operator such that the previous white noise assumption holds.

Each pair of antennas provides us with one visibility, which is a noisy Fourier component of the intensity image. Using an array of n radio antennas allows us to sample $\binom{n}{2} = (n^2 - n)/2$ points in the uv -plane (or Fourier plane). The distribution of these points depends on the configuration of the radio antenna array. If different time intervals are considered, the Earth's rotation can be exploited

to increase the number of uv points. The uv coverage is incomplete in all practical cases, and the measurements are noisy. Therefore, the linear operator Φ is ill-posed. If we also consider a large number of measurements, recovering \mathbf{x} from \mathbf{y} becomes a challenging inverse problem.

The most basic reconstruction of \mathbf{x} is often referred to as the naturally weighted dirty image that we will denote from now on as $\hat{\mathbf{x}}_{\text{dirty}}$ and call by dirty image as calibration weights are not currently being taken into account in this work. This estimation is obtained by applying the adjoint of Φ to the visibilities \mathbf{y} . To obtain a higher fidelity solution to the RI imaging inverse problem, we must regularize the problem by incorporating some prior information about the desired solutions \mathbf{x} . A broad range of methods can be characterized by what type of prior information is used to regularize the inverse problem and which algorithm is used to compute the reconstructed image $\hat{\mathbf{x}}$.

2.2 Sparsity-based regularization

The last two decades have brought us a great number of RI imaging methods based on sparse representations. The prior information exploited is that the solution \mathbf{x} is known to be sparsely represented in some bases or dictionaries. The bases are often built using multiscale wavelets, or a dictionary is constructed with a collection of wavelets (Mallat 2008). We can represent our image \mathbf{x} in a dictionary $\Psi \in \mathbb{C}^{N \times L}$,

$$\mathbf{x} = \Psi \mathbf{a} = \sum_{i=1}^L \Psi_i a_i, \quad (4)$$

where $\mathbf{a} \in \mathbb{C}^L$ is a vector of coefficients of \mathbf{x} weighting the corresponding dictionary atoms of Ψ . The assumption made to regularize the inverse problem is that \mathbf{a} is sparse or compressible, meaning that most of the coefficients are zero-valued or near zero, respectively. An array \mathbf{a} is called k -sparse if it has only k non-zero elements, which can be written as $\|\mathbf{a}\|_0 = k$, where $\|\cdot\|_0$ denotes the ℓ_0 pseudo-norm.

Sparsity should be ideally enforced through the ℓ_0 pseudo-norm, which is non-convex. Consequently, a convex relaxation to the ℓ_1 norm is used, which is a sparsity-promoting norm. The optimization problem is formulated such that its solution coincides with the inverse problem solution. Therefore, the inverse problem can be tackled with an optimization algorithm. The optimization objective comprises two competing terms: (i) a data-fidelity term $f(\cdot)$ that promotes consistency with the observed visibilities and depends on the statistics of the noise \mathbf{n} ; and (ii) a regularization term $r(\cdot)$ that encodes our prior knowledge of \mathbf{x} . The optimization problem reads

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\operatorname{argmin}} f(\mathbf{x}) + \sum_k \lambda_k r_k(\mathbf{x}), \quad (5)$$

where we are using a sum of regularization terms r_k , each with its corresponding regularization strength parameter λ_k . Substituting the RI data fidelity and sparsity-enforcing regularization in an overdetermined wavelet dictionary Ψ terms into equation (3) we obtain

$$\hat{\mathbf{x}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\operatorname{argmin}} \frac{1}{2\sigma^2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \lambda \|\Psi^\dagger \mathbf{x}\|_1, \quad (6)$$

where σ is the noise standard deviation.

The previous formulation is referred to as unconstrained. Other works consider the constrained formulation, which minimizes the ℓ_1 term with respect to a hard ℓ_2 -ball constraint over \mathbf{y} with a radius of ϵ , which is related to the noise's σ (Carrillo et al. 2012; Pratley et al. 2018). In this article, we will focus on the unconstrained formulation

as it has a natural Bayesian interpretation. Obtaining the solution from equation (6) involves solving a convex optimization problem, where we have the sum of a differentiable and a non-differentiable term. Proximal algorithms (Parikh & Boyd 2014) are well suited to tackle such optimization problems. Recent developments brought us a wide collection of proximal optimization algorithms, such as the forward-backward (FB) algorithm (Combettes & Pesquet 2009), the FISTA algorithm (Beck & Teboulle 2009), the alternating direction method of multipliers (ADMM; Boyd et al. 2011), and the primal-dual forward-backward algorithm (Chambolle & Pock 2011; Condat 2013), to mention a few. A rich literature exists exploiting the aforementioned concepts to tackle the RI imaging problem (Wiaux et al. 2009; McEwen & Wiaux 2011; Carrillo et al. 2012, 2014; Onose et al. 2016; Pratley et al. 2018, 2019a, b; Cai et al. 2018b; Pratley & McEwen 2019). For example, the sparsity averaging reweighted analysis (SARA) family of methods (Carrillo et al. 2012) use an overcomplete dictionary composed of a concatenation of the Dirac basis and the first eight Daubechies wavelets (Daubechies 1992) and has shown good performance in RI imaging.

2.3 CLEAN

Precursor of RI image reconstructions, the CLEAN algorithm (Högbom 1974) is a highly successful RI imaging method and it is still being used (Event Horizon Telescope Collaboration 2019a, b) despite various negative characteristics. The CLEAN algorithm is a non-linear iterative method that assumes a sparse sky model. CLEAN iteratively removes the contribution of the brightest source convolved with the instrument's point spread function or dirty beam. This method can be interpreted as a matching pursuit algorithm (Wiaux et al. 2009), or an ℓ_0 regularization with a basis composed of a sum of Dirac spikes. Several extensions of CLEAN have been developed over time (Bhatnagar & Cornwell 2004; Cornwell 2008; Stewart, Fenech & Muxlow 2011; Offringa et al. 2014; Offringa & Smirnov 2017) achieving better reconstruction performance. See Rau et al. (2009) for a review of CLEAN-based algorithms.

On top of a very early introduction, the success of CLEAN resides in its *scalability*. However, CLEAN has been shown to produce artefacts when point sources do not well describe the underlying sky model limiting CLEAN's image quality and justifying the need for more advanced techniques, e.g. based on Section 2.2. CLEAN often requires manual intervention, making its use less practical. Furthermore, CLEAN and its extensions do not provide meaningful uncertainty quantification of its reconstruction.

2.4 Learned approaches

The advent of deep learning models has affected many imaging applications, and RI imaging is no exception. Handcrafted models and priors are limited in the information they can capture or represent with respect to recent, more expressive neural networks. Learned or data-driven methods can encode complex information existing in the data, e.g. astrophysical simulations, used in their training. In general, these approaches produce reconstructions with improved quality, a computational speed-up, or both. These reasons make learned approaches very relevant to the RI imaging reconstruction problem. However, there are issues regarding the robustness of learned methods to data distribution shifts (Hendrycks et al. 2020) and scalable methods for uncertainty quantification to the reconstruction.

Allam (2016) proposed a learned method for RI imaging based on convolutional neural networks (Dong et al. 2016) originally considered for super-resolution. The approach consists of learning

to post-process dirty images with variants for both, known and unknown PSF. More recently, Gheller & Vazza (2021) proposed to use a convolutional denoising autoencoder to learn to post-process radio images, e.g. the dirty image or CLEAN's output. Connor et al. (2022) proposed a residual deep neural network (DNN) coined POLISH that works as a learned post-processing and super-resolution network. The DNN is based on the architecture proposed in Yu et al. (2018) and takes as input dirty images at different wavelengths and resolutions. POLISH outputs a clean image at a higher resolution for each wavelength and shows a better reconstruction quality than CLEAN. The proposed method has been applied to simulations from the upcoming Deep Synoptic Array-2000 (Hallinan et al. 2019) and real data from the very large array (VLA; Perley et al. 2011).

The Plug-and-Play (PnP) framework (Venkatakrishnan, Bouman & Wohlberg 2013) provides a way to incorporate a deep learning model into a modern optimization algorithm. The central idea is to replace a proximal regularization term with a denoising deep neural network. Ryu et al. (2019) studied conditions for the convergence of PnP algorithms. Pesquet et al. (2021) proposed a new term for the denoiser's training loss that enforces the firm non-expansiveness of the denoiser, which is usually deep learning-based. This training procedure allows the denoiser to suit a PnP framework with theoretical convergence conditions. The PnP framework with the non-expansiveness enforced to the deep learning-based denoiser has been applied to the RI imaging problem in Terris et al. (2022), where the approach has been called AIRI for Artificial Intelligence for Regularisation in RI imaging. The approach achieved similar or better performance than competing prior-based approaches whilst validating a significant acceleration potential. The AIRI method was later validated on observations from the Australian Square Kilometre Array Pathfinder (ASKAP; Wilber et al. 2023).

Two learned approaches for an interferometric-based imager named segmented planar imaging detector for electro-optical reconnaissance (SPIDER) were proposed by Mars et al. (2023). The first approach consists of a learned post-processing step from the dirty reconstruction based on a convolutional U-Net architecture (Ronneberger, Fischer & Brox 2015). The second approach consists of a learned multiscale iterative method coined GU-Net, which incorporates the measurement operator to include measurement information at the different steps and scales of the method. GU-Net is more efficient than standard unrolling methods due to its multiscale nature. The numerical results show an improved reconstruction quality and a faster convergence than proximal optimization-based methods. In the following work (Mars et al. 2024), the GU-Net was applied to the RI imaging problem. The variations of the uv -coverage are handled by training the neural network on a broad distribution of simulated uv -coverages and subsequently fine-tuning the network for a specific sampling distribution.

Aghabiglou et al. (2023) recently proposed a series of DNNs that combines notions of PnP algorithms and unrolled optimization methods (Adler & Öktem 2018; Monga, Li & Eldar 2019). Each DNN is trained to transform a back-projected residual into an image residual, thus ideally improving the reconstruction of the previous iteration. The results show a significant speed-up with respect to AIRI or SARA-based methods while maintaining a similar reconstruction quality. Other recent approaches based on deep neural networks include: Wang et al. (2023) who proposed a denoising diffusion probabilistic model conditioned on the visibilities and the dirty reconstruction; and Schmidt et al. (2022) who proposed a convolutional neural network based on residual blocks that intend to inpaint the measurements, or recover the entire uv plane from an incomplete coverage.

2.5 Bayesian framework

Bayesian inference provides a principled statistical framework to solve the inverse problem in equation (3) with statistical guarantees. This framework builds upon Bayes' famous theorem,

$$p(\mathbf{x}|\mathbf{y}) = \frac{\overbrace{p(\mathbf{y}|\mathbf{x})}^{\text{Likelihood}} \overbrace{p(\mathbf{x})}^{\text{Prior}}}{\underbrace{\int_{\mathbb{R}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x})d\mathbf{x}}_{\text{Bayesian evidence}}}. \quad (7)$$

Bayes' theorem relates the posterior distribution to the likelihood and prior terms that are the main constituents of a Bayesian model. The likelihood is associated with the data-fidelity term depending on the observational model and the noise statistics. The prior models expected properties of the solution \mathbf{x} , for example, smoothness, and piecewise regularity. This prior knowledge regularizes the estimation problem.

The term in the denominator, commonly known as the Bayesian evidence, does not depend on \mathbf{x} as we are marginalizing over that variable, and it describes the likelihood of the observed data based on the modelling assumptions. The Bayesian evidence is crucial for making Bayesian model comparison (Robert 2007), which provides us with a consistent way to compare models. Such high dimensional integrals can be effectively estimated by, for example, nested sampling techniques (Skilling 2006; Ashton et al. 2022), or the recently introduced learned harmonic mean estimator (McEwen et al. 2021; Spurio Mancini et al. 2022; Polanska et al. 2023). Recent developments have focused on nested sampling to compute the model evidence in high-dimensional imaging problems with sparsity-based handcrafted priors (Cai, McEwen & Pereyra 2022) and deep learning-based priors (McEwen et al. 2023). Carrying out model selection is out of the scope of this work.

Under the Bayesian framework, we have the posterior distribution $p(\mathbf{x}|\mathbf{y})$ which assigns a probability to each possible solution \mathbf{x} given some observations \mathbf{y} and a model \mathcal{M} consisting of the likelihood and prior terms. In imaging settings, explaining the information contained in the posterior distribution is not trivial due to its high-dimensional nature. The posterior distribution is generally characterized by samples computed by MCMC sampling. Efficiently sampling from high-dimensional posterior distributions is a current research topic, see e.g. Klatzer et al. (2023). Once $p(\mathbf{x}|\mathbf{y})$ is defined, we can say that the reconstruction method will be a point estimator of the posterior that will provide us with $\hat{\mathbf{x}}$. There are several choices for point estimators (Robert 2007; Arridge et al. 2019), each with advantages and drawbacks. Some examples are a sample from the posterior, $\hat{\mathbf{x}} \sim p(\mathbf{x}|\mathbf{y})$, the maximum-a-posteriori estimator, $\hat{\mathbf{x}} = \operatorname{argmax}_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$, or the posterior mean, $\hat{\mathbf{x}} = \mathbb{E}[\mathbf{x}|\mathbf{y}]$.

The posterior also provides us with consistent ways of quantifying the uncertainty of the chosen point estimate or reconstruction (Robert 2007). For example, one way to represent uncertainty is to compute the posterior standard deviation. The pixels with a higher standard deviation are less constrained by the data and the prior allowing for more significant fluctuations.

One of the most significant drawbacks of Bayesian imaging methods is that they are known to be computationally expensive, even if there is a continuous effort targetting the scalability of these methods (Pereyra 2016; Durmus, Moulines & Pereyra 2018; Pereyra, Mieleś & Zygalakis 2020; Pereyra, Vargas-Mieleś & Zygalakis 2022; Klatzer et al. 2023).

2.5.1 Maximum-a-posteriori estimation

The MAP estimator is particularly interesting in high-dimensional problems like RI imaging as its formulation allows us to bypass the need for sampling from the posterior. Consequently, its computational footprint is significantly reduced. The likelihood and prior terms can be rewritten as $p(\mathbf{y}|\mathbf{x}) = \exp[-f(\mathbf{x}, \mathbf{y})]$ and $p(\mathbf{x}) = \exp[-g(\mathbf{x})]$, respectively. The functions f and g are the likelihood and prior potentials. Using Bayes' theorem in equation (7), we can rewrite the MAP estimation as follows

$$\hat{\mathbf{x}}_{\text{MAP}} = \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x}|\mathbf{y}) = \operatorname{argmax}_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \quad (8)$$

The previous optimization problem can be reformulated using the monotonicity of the logarithm as follows

$$\begin{aligned} \hat{\mathbf{x}}_{\text{MAP}} &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) \\ &= \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} f(\mathbf{x}, \mathbf{y}) + g(\mathbf{x}). \end{aligned} \quad (9)$$

One advantage of the MAP estimator is that equation (9) can be tackled efficiently with optimization algorithms. We refer the reader to Pereyra (2019) for a deeper analysis of MAP estimation.

Coming back to the RI imaging inverse problem from equation (3), we can define a (white) Gaussian likelihood,

$$p(\mathbf{y}|\mathbf{x}) \propto \exp \left[-\frac{1}{2\sigma^2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \right] \quad (10)$$

and a sparsity-inducing Laplace-type prior defined as

$$p(\mathbf{x}) \propto \exp \left[-\lambda \|\Psi^\dagger \mathbf{x}\|_1 \right]. \quad (11)$$

Upon substitution of equations (10) and (11) into equation (9), the MAP optimization problem coincides with the one in equation (6). Therefore, the MAP reconstruction, $\hat{\mathbf{x}}_{\text{MAP}}$, matches $\hat{\mathbf{x}}$ from equation (6). Hence, sparsity-based approaches are MAP estimations with a prior based on the sparsity-promoting ℓ_1 norm in a given dictionary, e.g. wavelets.

2.5.2 Uncertainty quantification: more than a point estimate

Computing a good reconstruction for an inverse problem in the form of equation (3) can itself be challenging. Moreover, the reconstruction is often insufficient for many scientific applications that require further quantification of the result. This demand opens the door to uncertainty quantification, which provides more than a point estimate. The Bayesian framework provides us with formidable tools to do uncertainty quantification. For example, if we choose the MAP estimator as our reconstruction following the model in Section 2.5.1, we obtain the same reconstruction as in Section 2.2, which is the solution of equation (6). However, with the Bayesian framework, we can sample from the posterior and estimate the posterior standard deviation, perform a Bayesian hypothesis test of some image structure (Cai et al. 2018a; Price et al. 2021b), or compute other pixel-wise uncertainty measurements like local credible intervals (LCI; Cai et al. 2018a; Price et al. 2019).

2.5.3 Bayesian inference via MCMC sampling

Recent developments (Durmus, Moulines & Pereyra 2022) have considerably reduced the computational complexity of sampling high-dimensional posterior distributions in imaging inverse problems. Proximal MCMC sampling algorithms (Pereyra 2016; Durmus et al. 2018) extend the class of posterior distributions that can be studied

by allowing the use of non-smooth terms. Sparse regularizers have been widely used in RI imaging (Carrillo et al. 2012, 2014; Pratley et al. 2018; Cai et al. 2018a), and are usually enforced through a non-smooth ℓ_1 term.

Let us note π the target probability distribution that we are interested in sampling from, which in our case will be the posterior $p(\mathbf{x}|\mathbf{y})$. We consider a Langevin diffusion process on \mathbb{R}^N such that its stationary distribution is π . Assuming that $\pi \in \mathcal{C}^1$ with Lipschitz gradients, we write the Langevin diffusion as the following stochastic process

$$dL(t) = \frac{1}{2} \nabla \log \pi [L(t)] dt + dW(t), \quad L(0) = l_0, \quad (12)$$

where W is a N -dimensional Brownian motion. A usual discrete-time approximation of the Langevin diffusion consists of a forward Euler approximation with a step size δ , known as the Euler-Maruyama approximation (Kloeden & Platen 2011). The resulting algorithm is known as the unadjusted Langevin algorithm (ULA),

$$\mathbf{l}^{(m+1)} = \mathbf{l}^{(m)} + \delta \nabla \log \pi [\mathbf{l}^{(m)}] + \sqrt{2\delta} \mathbf{w}^{(m+1)}, \quad (13)$$

where $\mathbf{w}^{(m+1)} \sim \mathcal{N}(0, \mathbf{I}_N)$ is the discrete counterpart of $W(t)$. The ULA-based Markov chain converges to π with an asymptotic bias due to discretization. The bias can be accounted for with a subsequent Metropolis-Hasting (MH) accept-reject step. Adding the MH step corrects the bias but increases the algorithm's computational complexity. The ULA algorithm with the subsequent MH step is known as the Metropolis-adjusted Langevin algorithm (MALA).

The ULA algorithm requires the target density π to be continuously differentiable with Lipschitz gradients. Let us now consider $\pi(\mathbf{x}) \propto \exp[-f(\mathbf{x}) - g(\mathbf{x})]$, where $f \in \mathcal{C}^1$ with Lipschitz gradient and g is non-smooth but is a lower semicontinuous convex function that admits a proximal operator (Parikh & Boyd 2014). Proximal MCMC algorithms (Pereyra 2016) relax this assumption by approximating g , a non-smooth term in π , by its Moreau-Yosida envelope g^γ . The Moreau-Yosida approximation satisfies

$$\nabla g^\gamma(\mathbf{x}) = \frac{1}{\gamma} (\mathbf{x} - \text{prox}_\gamma^g(\mathbf{x})), \quad (14)$$

$$\text{prox}_\gamma^g(\mathbf{x}) := \underset{\mathbf{u} \in \mathbb{R}^N}{\text{argmin}} \left\{ g(\mathbf{u}) + \frac{1}{2\gamma} \|\mathbf{u} - \mathbf{x}\|_2^2 \right\}, \quad (15)$$

where γ is the Moreau-Yosida approximation parameter and the proximal operator may or may not have a closed-form expression. Consequently, the non-smooth target density π is approximated by the smooth π_γ , which replaces the g term with its Moreau-Yosida approximation g^γ . The Markov chain targeting π_γ writes

$$\mathbf{l}^{(m+1)} = \left(1 - \frac{\delta}{\gamma} \right) \mathbf{l}^{(m)} + \frac{\delta}{\gamma} \text{prox}_\gamma^g(\mathbf{l}^{(m)}) - \delta \nabla f(\mathbf{l}^{(m)}) + \sqrt{2\delta} \mathbf{w}^{(m+1)}, \quad (16)$$

and it is known as Moreau-Yosida regularized ULA (MYULA). If we add an MH step targeting the non-differentiable distribution π , the MCMC algorithm is known as Proximal MALA (Px-MALA). The proximal MCMC algorithms previously mentioned can be further accelerated by replacing the Euler-Maruyama approximation with the more involved Runge-Kutta-Chebyshev approximation (Abdulle, Almuslimani & Vilmart 2018), giving rise to the SK-ROCK (Pereyra et al. 2020) algorithm.

Cai et al. (2018a) exploited the MYULA and Px-MALA algorithms to sample from the posterior in the RI imaging problem. The model is based on a Gaussian likelihood as in equation (10) and a sparsity promoting prior akin to equation (11). However, the

framework can be used with more complex noise models (Melidonis et al. 2023), e.g. Poisson noise. In Cai et al. (2018a), the RI image reconstruction relies on the minimum mean squared error (MMSE) estimator based on the posterior mean, while in Cai et al. (2018b), the MAP is considered.

3 SCALABLE BAYESIAN DATA-DRIVEN IMAGING WITH UNCERTAINTY QUANTIFICATION

QUANTIFAI,¹ a scalable Bayesian data-driven method with uncertainty quantification is motivated by three principles:

(1) *Scalability*: The RI imaging inverse problem demands scalability for a method to be useful in real astronomical data scenarios such as SKA. The most time-consuming operation is evaluating the measurement operator Φ in the likelihood function. It is, therefore, essential to minimize the number of likelihood evaluations. For these reasons, we limit ourselves to the MAP estimator for our reconstruction corresponding to the solution of a convex optimization problem which converges quickly. We need to avoid sampling-based approaches as they are prohibitively expensive in terms of computations.

(2) *High-quality reconstructions*: To improve the quality of our reconstruction, we consider data-driven or learned priors that can better encode the expected image structures. In Section 2.4, we have already seen that data-driven approaches can better represent complex imaging priors and provide reconstructions superior to handcrafted priors, such as sparsity-promoting priors based on wavelet dictionaries.

(3) *Uncertainty quantification*: There are many ways to quantify uncertainty based on sampling the posterior distribution. However, using sampling-based methods is prohibitively expensive, and one of our key criteria is computational scalability. Therefore, we need to restrict ourselves to log-concave posteriors, which is equivalent to saying that the addition of our potentials $f + g$ has to be convex, and to explicit potentials. As we will later describe in more detail in Section 4, the first restriction enables the use of efficient methods relying on the concentration of probability for high-dimensional log-concave distribution (Pereyra 2017). Consequently, we can use approximate posterior information bypassing sampling methods. These methods are orders of magnitude faster resulting in a scalable Bayesian UQ method. In a nutshell, we require the posterior potential to be convex and explicit for scalable UQ. The likelihood is typically convex for RI imaging problems so we will enforce the prior potential g to be convex and explicit. The requirement of explicit potentials will be explained in Section 4.

We continue by introducing the data-driven convex regularizers and the optimization algorithm used to compute the MAP estimation for the proposed method.

3.1 Learned convex regularizers

As stated before, we need an expressive regularizer that is convex and has an explicit potential. Modern regularizers based on deep neural networks, like convolutional neural networks, used in RI imaging reconstruction methods satisfy neither of the two constraints. This last constraint, i.e. with an explicit potential required by the UQ approach, excludes a range of denoisers whose potentials

¹Code available at github.com/astro-informatics/QuantifAI

are defined implicitly. PnP approaches (Terris et al. 2022) only require the denoising of the image without explicitly computing the regularization potential. For example, a typical iteration from a PnP algorithm writes

$$\mathbf{x}_{k+1} = \mathbf{D}(\mathbf{x}_k - \gamma \nabla f(\mathbf{x}_k)), \quad (\forall k \in \mathbb{N}), \quad (17)$$

where \mathbf{D} is the denoiser, f is the data-fidelity term, γ is the step size, and k is the iteration number. The algorithm's convergence can be assured if \mathbf{D} and the stepsize satisfies some conditions (Ryu et al. 2019; Pesquet et al. 2021). Even if the denoiser \mathbf{D} is convex, we cannot use it for our approach as we must evaluate the potential.

Mukherjee et al. (2020) proposed a learned convex regularizer parametrized by the architecture of a deep input-convex neural network (ICNN; Amos, Xu & Zico Kolter 2016), which is convex by construction. The training of the regularizer is done with an adversarial framework introduced by Lunz, Öktem & Schönlieb (2018).

Very recently, a learnable convex-ridge regularizer neural network (CRR-NN;² Goujon et al. 2023b) has been proposed, which comes with the required properties of being convex and having an explicit potential. In addition, the model focuses on being reliable and interpretable while still being expressive enough to provide excellent reconstruction quality. The CRR-NN regularizer, R_θ , has the form

$$R_\theta : \mathbb{R}^N \mapsto \mathbb{R}, \quad R_\theta(\mathbf{x}) = \sum_{i=1}^{N_C} \sum_k \psi_i((\mathbf{h}_i * \mathbf{x})[k]), \quad (18)$$

where \mathbf{h}_n are learnable 2D convolution kernels, $(\mathbf{h}_i * \mathbf{x})[k]$ denotes the k -th pixel of the resulting convolution, N_C is the number of channels or kernels, $\psi_i : \mathbb{R} \mapsto \mathbb{R}$ are learnable non-linear convex profile functions with a Lipschitz continuous derivative, i.e. $\psi_i \in C^{1,1}(\mathbb{R})$, and θ in R_θ represents all learnable parameters. The convexity constraint on the learnable activation functions, ψ_i , is enforced by making the pointwise $\sigma_i : \mathbb{R} \rightarrow \mathbb{R}$ monotonically increasing, with $\psi'_i = \sigma_i$, where $\sigma_i \in C_{\uparrow}^{0,1}(\mathbb{R})$, and $C_{\uparrow}^{0,1}(\mathbb{R})$ is the set of scalar Lipschitz continuous and increasing functions on \mathbb{R} . The σ_i functions are chosen as learnable linear splines, which have been shown to be more expressive than ReLU functions in Propositions 3.3 and 3.5 from Goujon et al. (2023b, section III.B). The main difference between a prior based on the CRR-NN and a wavelet dictionary is that the kernels (or filters) and the activation (or thresholding) functions are learnt in the first one. In the second one, they are fixed or handcrafted. We refer the reader to Goujon et al. (2023b, figs 5 and 6) for examples of learned kernels \mathbf{h}_n and activation functions for two trainings with two different noise levels. See Bohra et al. (2020) and Goujon et al. (2023b), for more information on learnable splines.

In the spirit of PnP approaches, the CRR-NN training is based on the denoising problem that reads

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{x} - \mathbf{y}\|_2^2 + \lambda R_\theta(\mathbf{x}), \quad (19)$$

where \mathbf{y} is a noisy version of \mathbf{x} , and λ is a parameter controlling the regularization strength. The denoising problem is addressed through the fixed point of the problem, which given the convexity assumptions, is unique. A gradient step of equation (19) reads

$$T_{R_\theta, \lambda, \alpha}(\mathbf{x}) = \mathbf{x} - \alpha((\mathbf{x} - \mathbf{y}) + \lambda \nabla R_\theta(\mathbf{x})), \quad (20)$$

where α is the stepsize. Convergence can be guaranteed if the stepsize satisfies $\alpha \in (0, 2/(1 + \lambda \operatorname{Lip}(\nabla R_\theta)))$, where $\operatorname{Lip}(\cdot)$ denotes

the Lipschitz constant. By composing t gradient descent updates of equation (20), i.e. a t -fold composition, we obtain a multigradient step denoiser that we denote $T_{R_\theta, \lambda, \alpha}^t$ following the notation of Goujon et al. (2023b).

The denoising problem in equation (19) can be formulated as a fix point problem for the t -step denoiser $T_{R_\theta, \lambda, \alpha}^t$ as follows,

$$T_{R_\theta, \lambda, \alpha}^t(\mathbf{y}) \approx \mathbf{x}. \quad (21)$$

We build the CRR-NN training by penalizing the residual of the fix point problem in equation (21) with a loss function \mathcal{L} , for a training set of pairs of noiseless and noisy images $\{\mathbf{x}^{(m)}, \mathbf{y}^{(m)}\}_{m=1}^M$, and reads

$$\theta_i^*, \lambda_i^* \in \operatorname{argmin}_{\theta, \lambda} \sum_{m=1}^M \mathcal{L}(T_{R_\theta, \lambda, \alpha}^t(\mathbf{y}^{(m)}), \mathbf{x}^{(m)}). \quad (22)$$

After having trained the denoiser, we define our prior potential as

$$g(\mathbf{x}) = \frac{\lambda}{\mu} R_\theta(\mu \mathbf{x}), \quad (23)$$

where we have dropped the θ_i^*, λ_i^* notation for θ, λ and added a scaling parameter, μ , to boost performance following Goujon et al. (2023b). For the optimization algorithm, we need the Lipschitz constant of the gradient of the potential in equation (23), which can be expressed as

$$\operatorname{Lip}(\nabla g) = \lambda \mu \operatorname{Lip}(\nabla R_\theta) \leq \lambda \mu \|\mathbf{W}^T \Sigma_\infty \mathbf{W}\|, \quad (24)$$

which is calculated in Goujon et al. (2023b, Prop. IV.1), and $\Sigma_\infty = \operatorname{diag}(\|\sigma'_1\|_\infty, \dots, \|\sigma'_{N_C}\|_\infty)$, and $\mathbf{W} = [\mathbf{w}_1 \dots \mathbf{w}_{N_C}]^T$ where \mathbf{w}_i corresponds to the filter \mathbf{h}_i : $\mathbf{h}_i * \mathbf{x} = \mathbf{w}_i^T \mathbf{x}$.

3.2 Computing our reconstruction: the MAP

In our case, computing the MAP reduces to solving a convex optimization problem. Following equation (9), the optimization problem we address is the following one,

$$\hat{\mathbf{x}}_{\text{MAP}} = \operatorname{argmin}_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2\sigma^2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \frac{\lambda}{\mu} R_\theta(\mu \mathbf{x}) + \iota_{\mathbb{R}^N}(\mathbf{x}), \quad (25)$$

where in addition we include $\iota_{\mathbb{R}^N}$, an indicator function enforcing the reconstructed image to be real. The proximal operator of the indicator function to a convex set is known and it amounts to a projection onto that convex set. In the last term of equation (25) the proximal operator of the reality constraint is the projection of the vector to the real number, which is written as $\operatorname{Re}(\cdot)$. We have assumed a (white) Gaussian likelihood and the prior term is based on a previously trained CRR-NN. The CRR-NN is smooth with Lipschitz continuous gradients. However, the non-smoothness of the reality enforcing constraint forces us to rely on proximal algorithms (Parikh & Boyd 2014) instead of an accelerated gradient descent method (Nesterov 2018). In this case, we use the FISTA algorithm (Beck & Teboulle 2009).

For the optimization, we need the gradient of the likelihood and prior terms

$$\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) = \frac{1}{\sigma^2} (\Phi^\dagger (\Phi \mathbf{x} - \mathbf{y})), \quad (26)$$

$$\nabla g(\mathbf{x}) = \lambda \nabla R_\theta(\mu \mathbf{x}), \quad (27)$$

where, in our case, $(\cdot)^\dagger$ is the complex conjugate transpose.

To ensure the algorithm's convergence we use the stepsize $\tau = 1/L$, where $L = \operatorname{Lip}(\nabla_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}) + \nabla g(\mathbf{x}))$. We can estimate a simple

²https://github.com/axgoujon/convex_ridge_regularizers

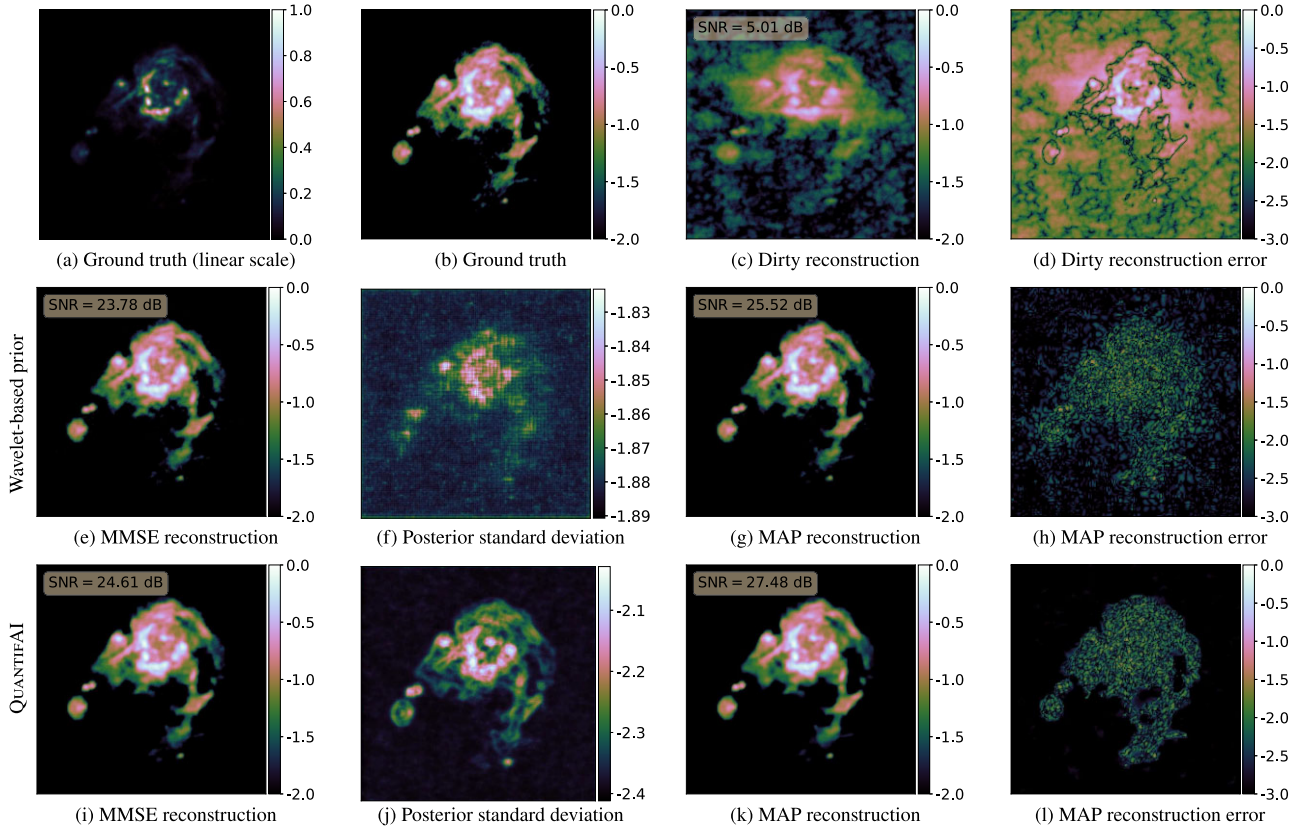


Figure 1. RI image reconstructions for M31. The images are shown in a \log_{10} scale except for subfigure (a). Top row: The first two images show the ground truth intensity image in linear and \log_{10} scales, respectively. The third image shows the dirty reconstruction, computed by applying the pseudo-inverse of the measurement operator on the observations. The fourth image shows the error of the dirty reconstruction with respect to the ground truth. Middle row: We show the results of the wavelet-based model. The first and second columns show the minimum mean squared error (MMSE) estimator and the posterior standard deviation, respectively. Both images are computed using the 5×10^4 generated posterior samples. The third column shows the MAP reconstruction obtained through an optimization algorithm. The fourth column depicts the error of the MAP reconstruction with respect to the ground truth. Bottom row: We present the results of the QUANTIFAI model. The columns are presented in the same order as for the wavelet reconstructions in the middle row. For every reconstruction, we display the SNR with respect to the ground truth in the top left corner. Compared with the wavelet-based model, QUANTIFAI recovers a reconstruction with a higher SNR and shows more meaningful uncertainties, which can be seen by comparing the posterior standard deviation and the MAP reconstruction error.

bound for the Lipschitz constant as follows

$$L \leq \text{Lip}(\nabla_x f(\mathbf{x}, \mathbf{y})) + \text{Lip}(\nabla g(\mathbf{x})) = L_{\text{likelihood}} + L_{\text{prior-CRR-NN}},$$

$$\leq \frac{\|\Phi^\dagger \Phi\|}{\sigma^2} + \lambda \mu \|\mathbf{W}^T \Sigma_\infty \mathbf{W}\|, \quad (28)$$

where we have exploited the result from equation (24), and $\|\Phi^\dagger \Phi\|$ denotes the spectral norm, which in the case of a linear operator coincides with its maximum singular value. In the simplified problem we are considering in Section 2.1 with gridded visibilities, we have that $\|\Phi^\dagger \Phi\| = 1$. If a more realistic linear operator should be considered, the maximum singular value could be computed iteratively via the power method (Golub & van Loan 2013).

We initialize the optimization with the naturally weighted dirty image, $\mathbf{x}_{(0)} = \text{Re}(\Phi^\dagger \mathbf{y})$. The optimization procedure is summarized in Algorithm 1. We optimize for a fixed number of iterations N_{max} , or until a tolerance criterion of ξ is reached. The stepsize is computed using the bound from equation (28).

4 SCALABLE UNCERTAINTY QUANTIFICATION

Enforcing the posterior's convexity and explicit potential opens the door to scalable UQ methodology that was unreachable otherwise.

Algorithm 1: FISTA (Beck & Teboulle 2009) tackling (25)

- 1 **Input:** $R_\theta, \Phi, \sigma, \mu, \lambda, \xi, a_{(1)} = 1, \mathbf{z}_{(1)} = \mathbf{x}_{(0)} = \text{Re}(\Phi^\dagger \mathbf{y}), \tau = 0.98/L$.
- 2 **Output:** $\hat{\mathbf{x}}_{\text{MAP}}$
- 3 **for** $n = 1, \dots, N_{\text{max}}$ **do**
- 4 $\mathbf{x}_{(n)} = \mathbf{z}_{(n)} - \tau \left(\frac{1}{\sigma^2} \text{Re}(\Phi^\dagger (\Phi \mathbf{z}_{(n)} - \mathbf{y})) + \lambda \nabla R_\theta(\mu \mathbf{z}_{(n)}) \right)$
- 5 $a_{(n+1)} = \frac{1}{2} (1 + \sqrt{4a_{(n)}^2 + 1})$
- 6 $\mathbf{z}_{(n+1)} = \mathbf{x}_{(n)} + \frac{a_{(n)} - 1}{a_{(n+1)}} (\mathbf{x}_{(n)} - \mathbf{x}_{(n-1)})$
- 7 **if** $\frac{\|\mathbf{x}_{(n)} - \mathbf{x}_{(n-1)}\|}{\|\mathbf{x}_{(n-1)}\|} < \xi$ **then**
- 8 **break**
- 9 **end**
- 10 **end**
- 11 set $\hat{\mathbf{x}}_{\text{MAP}} = \mathbf{x}_{(n)}$

The restriction to log-concave posteriors is the price we pay to gain great scalability. Our approach is based on the work from Pereyra (2017), which exploits concentration phenomena occurring in high-dimensional log-concave posteriors. The Bayesian high-posterior-density region can be approximated in log-concave models

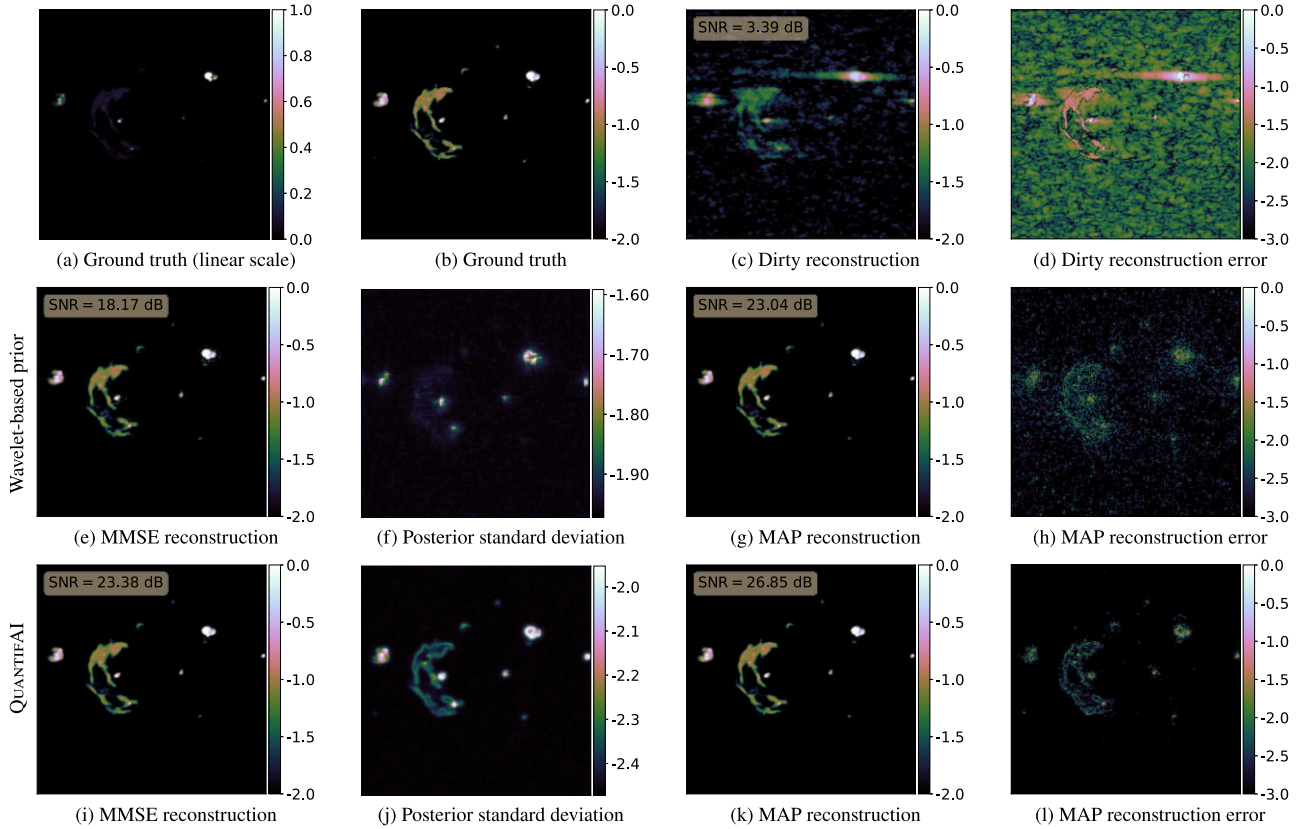


Figure 2. RI image reconstructions for W28. The order of the images follows the M31 results presented in Fig. 1. Compared with the wavelet-based model, QUANTIFAI recovers a reconstruction with a higher SNR and shows more meaningful uncertainties, which can be seen by comparing the posterior standard deviation and the MAP reconstruction error.

as the posterior probability mass tends to concentrate in particular regions on the parameter space. The approximation requires the MAP estimation, $\hat{\mathbf{x}}_{\text{MAP}}$, which we have already computed as it is the chosen point estimate for our reconstruction. This result allows us to estimate information from the posterior probability density function without MCMC sampling. In this section, we introduce the main result we exploit for UQ. We then describe the proposed scalable UQ methods and how to validate our results with Langevin-based MCMC sampling algorithms.

4.1 Highest posterior density regions

Let us define a posterior credible region with a credible level of $100(1 - \alpha)$ per cent as a set $C_\alpha \in \mathbb{R}^N$ satisfying

$$p(\mathbf{x} \in C_\alpha | \mathbf{y}) = \int_{\mathbf{x} \in \mathbb{R}^N} p(\mathbf{x} | \mathbf{y}) \mathbb{1}_{C_\alpha}(\mathbf{x}) d\mathbf{x} = 1 - \alpha, \quad (29)$$

with $\mathbb{1}_{C_\alpha}$ being unity if $\mathbf{x} \in C_\alpha$ and zero otherwise. There are many regions satisfying the previous equation. We will focus on the highest posterior density region (HPD), which is defined as

$$C_\alpha := \{ \mathbf{x} \in \mathbb{R}^N : f(\mathbf{x}) + g(\mathbf{x}) \leq \gamma_\alpha \}, \quad (30)$$

where f and g are the potentials of our likelihood and prior terms, and γ_α is a constant that defines a level-set of the log-posterior such that equation (29) holds. The HPD region has the property of having minimum volume (Section 5.5; Robert 2007).

Our posterior $p(\mathbf{x} | \mathbf{y}) = \exp[-f(\mathbf{x}) - g(\mathbf{x})]/Z$ is log-concave on \mathbb{R}^N , where Z is the Bayesian evidence. Then, following Pereyra

(2017, Theorem 3.1), for any $\alpha \in (4 \exp[-N/3], 1)$, the HPD region C_α from equation (30) is contained in

$$\hat{C}_\alpha = \{ \mathbf{x} \in \mathbb{R}^N : f(\mathbf{x}) + g(\mathbf{x}) \leq \hat{\gamma}_\alpha \}, \quad (31)$$

where

$$\hat{\gamma}_\alpha = f(\hat{\mathbf{x}}_{\text{MAP}}) + g(\hat{\mathbf{x}}_{\text{MAP}}) + \sqrt{N} \tau_\alpha + N, \quad (32)$$

with a positive constant $\tau_\alpha = \sqrt{16 \log(3/\alpha)}$ independent of $p(\mathbf{x} | \mathbf{y})$.

Theorem 3.2 in Pereyra (2017) gives the error analysis of the approximation, and we see that $0 \leq \hat{\gamma}_\alpha - \gamma_\alpha \leq \tau_\alpha \sqrt{N} + N$. Therefore, the error upper bound grows linearly with the dimension N , making \hat{C}_α a stable approximation of C_α . The error lower bound along with the convexity of $f + g$ guarantees the inclusion $C \subseteq \hat{C}$ and consequently the resulting approximation is a conservative one where the errors are overestimated.

After showing the main result allowing us to do UQ bypassing posterior sampling methods, it is clear from where the constraints of the prior come. The convexity is required to guarantee a log-concave posterior, as the likelihood potential is convex. The prior potential g needs to be explicit to compute the approximate HPD region using equation (32).

4.2 MAP-based UQ methods

We now explore different scalable UQ schemes based on the fast approximate implicit representation of the HPD region. For all the methods presented, we assume that we have already computed the $\hat{\mathbf{x}}_{\text{MAP}}$ estimation and the approximated HPD region threshold, $\hat{\gamma}_\alpha$.

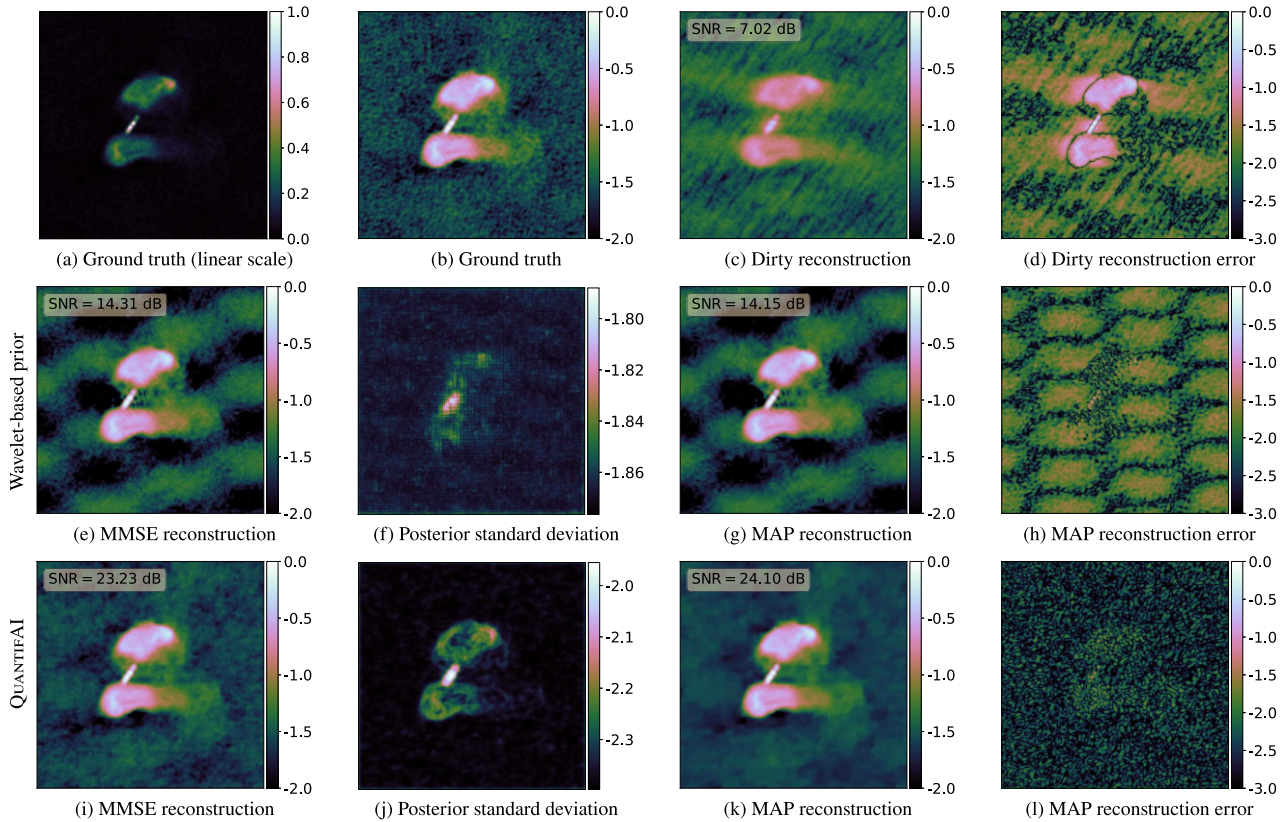


Figure 3. RI image reconstructions for 3C288. The order of the images follows the M31 results presented in Fig. 1. Compared with the wavelet-based model, QUANTIFAI recovers a reconstruction with a higher SNR and shows more meaningful uncertainties, which can be seen by comparing the posterior standard deviation and the MAP reconstruction error.

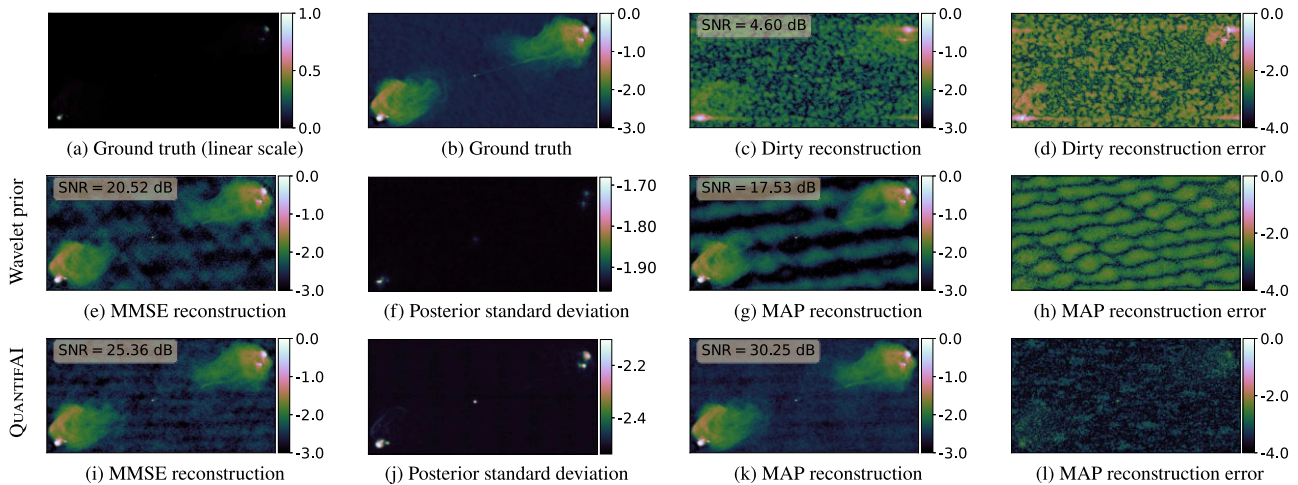


Figure 4. RI image reconstructions for Cygnus A. The order of the images follows the M31 results presented in Fig. 1. Compared with the wavelet-based model, QUANTIFAI recovers a reconstruction with a higher SNR and shows more meaningful uncertainties, which can be seen by comparing the posterior standard deviation and the MAP reconstruction error.

4.2.1 Bayesian hypothesis testing of structure

A useful UQ tool is to perform a *knock-out* hypothesis test to assess if a surrogate image still belongs to the HPD region (Cai et al. 2018a, b; Price et al. 2021b). First, the surrogate image \mathbf{x}_{sgt} is constructed by modifying the reconstruction, $\hat{\mathbf{x}}_{\text{MAP}}$. Then, it suffices to check if

$$f(\mathbf{x}_{\text{sgt}}) + g(\mathbf{x}_{\text{sgt}}) \leq \hat{\gamma}_\alpha. \tag{33}$$

If the inequality is satisfied, we cannot draw conclusions on the test we made, as \mathbf{x}_{sgt} still belongs to the HPD region. However, if the inequality does not hold, we can conclude that \mathbf{x}_{sgt} is out from the HDP region with a $100(1 - \alpha)$ per cent confidence level.

This test can answer a variety of questions about our reconstructed image. One example is to interrogate some structure in the image to see if it is a reconstruction artefact or is physically motivated. For this question, the surrogate image would be composed of

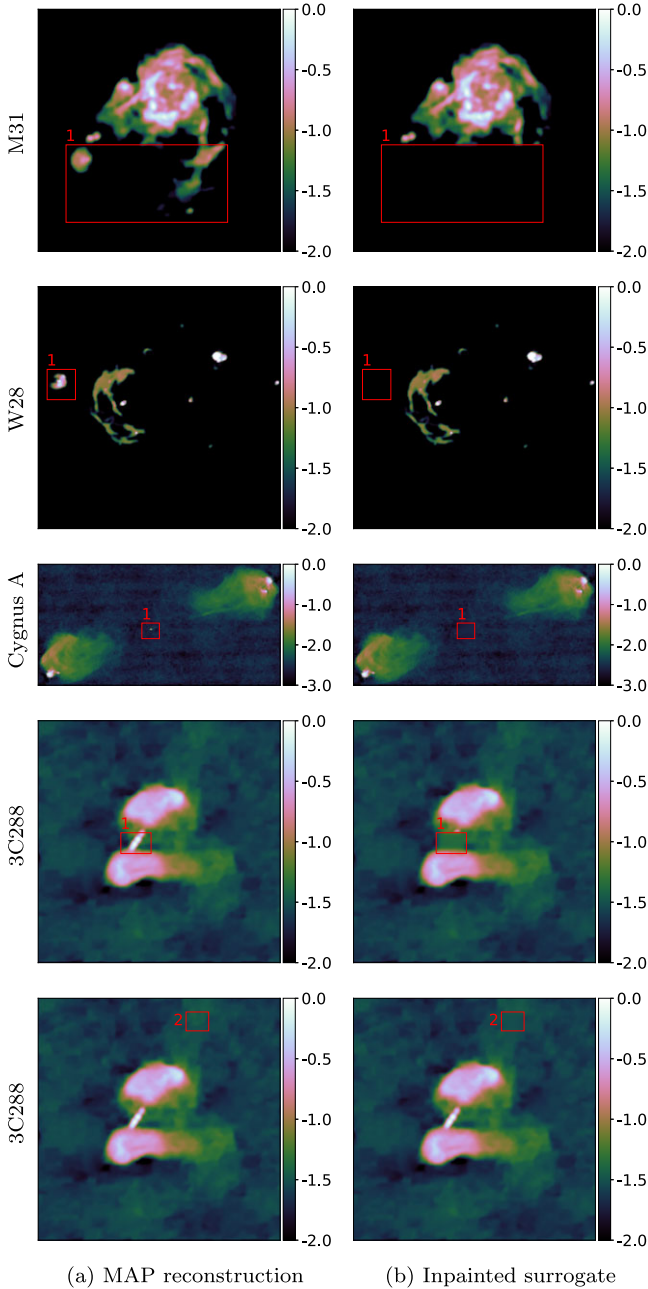


Figure 5. Hypothesis test of different regions of the four QUANTIFAI MAP reconstructions for M31, W28, Cygnus A, and 3C288. All the images are shown in \log_{10} scale. The left column shows the respective MAP reconstruction with the region of interest framed in an overlaid rectangle. The right column shows the surrogate images inpainted using the QUANTIFAI prior. The first four rows show regions corresponding to physical structures present in the ground truth images. The last row corresponds to a non-physical region. Results of the hypothesis tests are summarized in Table 2.

an image with the region of interest artificially inpainted with surrounding information. We need to take the inpainted image as our surrogate and evaluate equation (33) to see if the test is conclusive.

The image inpainting algorithm is built similarly as in Cai et al. (2018b) but adapted to the CRR-NN-based prior. We start by selecting a region of interest Ω_D , which is a subset of (typically contiguous) pixels from the image, where $\Omega_D \subseteq \Omega$, where Ω denotes

the set of all the image pixels. The region Ω_D will be inpainted with background information. We then inpaint this region iteratively minimizing R_θ based on the following scheme

$$\mathbf{x}_{\text{sgt.}(m+1)} = \hat{\mathbf{x}}_{\text{MAP}} \mathbb{1}_{\Omega - \Omega_D} + (\mathbf{x}_{\text{sgt.}(m)} - \alpha \lambda \nabla R_\theta(\mu \mathbf{x}_{\text{sgt.}(m)})) \mathbb{1}_{\Omega_D}, \quad (34)$$

where $\mathbb{1}$ are indicator functions, and $\mathbb{1}_{\Omega - \Omega_D}$ is a shorthand for $\mathbb{1}_\Omega - \mathbb{1}_{\Omega_D}$. We carry out a gradient step with the CRR-NN on the surrogate image and only update the region of interest. The hyperparameters, α , λ , and μ are set as in Algorithm 1.

Alternatively, Repetti, Pereyra & Wiaux (2019) presented a more sophisticated method to perform hypothesis testing of structure, which also exploits the approximations in equations (31)–(32). The method is dubbed Bayesian uncertainty quantification by optimization (BUQO), and to answer the hypothesis test, it aims to study the intersection of two sets. The first one is defined in equation (31) that corresponds to the MAP estimate. The second one describes the set of feasible inpainted images given a region of interest and a set of constraints of desired properties. If the set intersection is empty, the structure of interest is considered present in the image with confidence α from equation (31). Tang & Repetti (2023) proposed an extension of the BUQO method to inpaint with data-driven models. However, these methods involve solving an expensive optimization problem that does not scale with the high-dimensional settings we are considering in this work.

Another example is to interrogate the reconstruction to see if the fine structure of the image is physical or likely an artefact. To construct the surrogate image we convolve the region of interest, Ω_D , with a Gaussian smoothing kernel $G(0, \Sigma)$,

$$\mathbf{x}_{\text{sgt}} = \hat{\mathbf{x}}_{\text{MAP}} \mathbb{1}_{\Omega - \Omega_D} + (\hat{\mathbf{x}}_{\text{MAP}} * G(0, \Sigma)) \mathbb{1}_{\Omega_D}, \quad (35)$$

where $*$ denotes the 2D convolution operation and test equation (33).

4.2.2 Local credible intervals

Local credible intervals (LCIs) provide a tool to quantify spatial uncertainty per pixel at different resolutions. The LCIs are interpreted as Bayesian error bars for each pixel or superpixel, where with superpixel, we refer to a group of contiguous pixels. Cai et al. (2018a) computed LCIs using MCMC methods and then extended it in Cai et al. (2018b) to compute them based on the approximated HPD region based on the MAP. Price et al. (2019) also exploited MAP-based LCIs in another imaging inverse problem, mass-mapping, for weak gravitational lensing convergence reconstruction.

Let us write $\Omega = \{\Omega_i\}_{i=1}^M$ the set of superpixels that partition the domain of \mathbf{x} . This partition is such that $\Omega_i \cap \Omega_j = \emptyset, \forall i \neq j$ and $\Omega = \cup_i \Omega_i$. We denote the indicator of the superpixel Ω_i as ζ_{Ω_i} , that is one if the pixel belongs to the superpixel Ω_i and zero otherwise. The use of smaller or bigger superpixel sizes, i.e. $\|\zeta_{\Omega_i}\|_0$, allows us to visualize the LCIs at different scales. The calculation of the LCIs is based on computing an upper and lower bound for each superpixel. Each bound is defined by the constant value we need to add or subtract to the superpixel region so that the modified image exits the approximate HPD credible region \hat{C}_α . In other words, we compute the values that saturate the HPD region for each superpixel.

We can isolate the superpixel region by defining the following surrogate image

$$\mathbf{x}_{i,\xi} = \hat{\mathbf{x}}_{\text{MAP}} (\mathbf{I} - \zeta_{\Omega_i}) + (\xi + \bar{\mathbf{x}}_{\text{MAP},\Omega_i}) \zeta_{\Omega_i}, \quad (36)$$

where $\bar{\mathbf{x}}_{\text{MAP},\Omega_i}$ corresponds to the mean value of the pixels in the superpixel Ω_i , and $\xi \in \mathbb{R}$. We vary the superpixel value from its

mean by a uniform value ξ . The bounds for a superpixel Ω_i are computed as

$$\xi_{+, \Omega_i} = \max_{\xi} \{ \xi \mid f(\mathbf{x}_{i, \xi}, \mathbf{y}) + g(\mathbf{x}_{i, \xi}) \leq \hat{\gamma}_\alpha, \xi \in [0, +\infty) \}, \quad (37)$$

$$\xi_{-, \Omega_i} = \min_{\xi} \{ \xi \mid f(\mathbf{x}_{i, \xi}, \mathbf{y}) + g(\mathbf{x}_{i, \xi}) \leq \hat{\gamma}_\alpha, \xi \in (-\infty, 0] \}, \quad (38)$$

where we use the threshold $\hat{\gamma}_\alpha$ defined in equation (32). The calculation of each bound can be recast as a problem of finding the zero of a function. The class of root-finding algorithms is well adapted for this root-finding problem, and, in practice, we use the bisection method (Burden & Faires 1989). Price, Pratley & McEwen (2021a) proposed a faster way to compute the superpixel bounds by exploiting linearity that we could use to further accelerate the computation of ξ_{+, Ω_i} and ξ_{-, Ω_i} .

Once the bounds have been computed, we can collate the results for all superpixels and use the length of the LCIs to visualize the reconstruction uncertainty. The length of the LCI for superpixel Ω_i is defined as $l_i = \xi_{+, \Omega_i} - \xi_{-, \Omega_i}$, which we can visualize as an uncertainty image

$$\xi = \sum_i (\xi_{+, \Omega_i} - \xi_{-, \Omega_i}) \zeta_{\Omega_i}. \quad (39)$$

The choice of using the mean on $\bar{\mathbf{x}}_{\text{MAP}, \Omega_i}$ for the region of the superpixel that will be studied constitutes a deviation from the original MAP reconstruction. We find it more physical to move the averaged superpixel rather than moving the original pixels belonging to the superpixel by a constant value. This choice constitutes another approximation to the proposed scheme that has already approximated the HPD region in equation (31). Using superpixels allows us to gain sensibility and computing time at the expense of lowering the resolution of the LCI map, which can be a sensible trade-off for very large images.

We will later validate the computed LCIs using the posterior samples obtained from computing the posterior standard deviation at different superpixel sizes. The method requires turning each posterior sample into an image with M superpixels. We set the value of the superpixel to the mean of the values of belonging pixels.

4.2.3 Fast pixel uncertainty quantification at different scales

The MAP-based LCIs described in the previous section are orders of magnitude faster than their MCMC-based counterparts (Cai et al. 2018a, b). Nevertheless, to compute the LCIs, we still need to evaluate the likelihood potential, f , several times for each superpixel. As mentioned, evaluating the likelihood potential is by far the most time-consuming operation. The fact that we need to evaluate f several times for each subpixel might make the LCIs impractical for SKA-scale problems.

To overcome this issue, we propose a new approach relying on wavelet decomposition of the MAP reconstruction that reads

$$\hat{\mathbf{x}}_{\text{MAP}} = \Psi \hat{\mathbf{a}}_{\text{MAP}} = \sum_{i=1}^L \Psi_i \hat{\mathbf{a}}_{\text{MAP}, i}, \quad (40)$$

with a wavelet dictionary Ψ . We define the hard thresholding operator for $\mathbf{a} \in \mathbb{C}^L$ with a threshold of ξ_{th} ,

$$S_{\text{hard}, \xi_{\text{th}}}(\mathbf{a}) = [S_{\text{hard}, \xi_{\text{th}}}(a_1), \dots, S_{\text{hard}, \xi_{\text{th}}}(a_L)]^T, \quad (41)$$

as the point-wise application of the following hard-thresholding function

$$S_{\text{hard}, \xi_{\text{th}}}(a_i) = \begin{cases} 0, & \text{if } |a_i| \leq \xi_{\text{th}}, \\ a_i, & \text{otherwise.} \end{cases} \quad (42)$$

Let $\hat{\xi}_{\text{th}}$ be the thresholded value for which the thresholded image saturate the HPD region as follows

$$\hat{\xi}_{\text{th}} = \max_{\xi_{\text{th}}} \{ \xi_{\text{th}} \mid f(\hat{\mathbf{x}}_{\text{MAP}, \xi_{\text{th}}}, \mathbf{y}) + g(\hat{\mathbf{x}}_{\text{MAP}, \xi_{\text{th}}}) \leq \hat{\gamma}_\alpha, \hat{\mathbf{x}}_{\text{MAP}, \xi_{\text{th}}} = \Psi S_{\text{hard}, \xi_{\text{th}}}(\hat{\mathbf{a}}_{\text{MAP}}), \xi_{\text{th}} \in [0, +\infty) \}. \quad (43)$$

We can compute the threshold bound with a root-finding method, as was the case for the LCIs. The advantage of this formulation is that we are solving only one root-finding problem as opposed to one per superpixel in the LCIs calculation. This change considerably reduces the number of likelihood evaluations and, therefore, the computational complexity of the UQ method.

By computing the difference between the MAP, $\hat{\mathbf{x}}_{\text{MAP}}$, and the thresholded surrogate, $\hat{\mathbf{x}}_{\text{MAP}, \hat{\xi}_{\text{th}}}$, we obtain an estimation of the solution's uncertainty and this can give us information about possible errors in the MAP. Furthermore, we can compare the MAP and the thresholded surrogate image to estimate errors as a function of scale, thus exposing the different structures of our reconstruction.

Let us consider our wavelet transformation as a multiscale transform of level $J + 1$ (Mallat 2008; Starck, Murtagh & Fadili 2010). We can rewrite equation (40) showcasing the multiscale coefficients as follows

$$\hat{\mathbf{x}}_{\text{MAP}} = \Psi \hat{\mathbf{a}}_{\text{MAP}} = \sum_{l=0}^J \Psi_l \hat{\mathbf{a}}_{\text{MAP}, l}, \quad (44)$$

where $\hat{\mathbf{a}}_{\text{MAP}, l}$ are the coefficients corresponding to the l -th level of decomposition, and the zeroth level corresponds to the coarse scale. We can now build thresholded surrogate images at different scales by replacing the MAP wavelet coefficients at level l from equation (44) with the coefficients of the thresholded surrogate image $\hat{\mathbf{x}}_{\text{MAP}, \hat{\xi}_{\text{th}}}$. Let us write the thresholded surrogate image at level j as follows

$$\hat{\mathbf{x}}_{\text{MAP}, \hat{\xi}_{\text{th}}, j} = \sum_{\substack{l=0, \\ l \neq j}}^J \Psi_l \hat{\mathbf{a}}_{\text{MAP}, l} + \Psi_j \hat{\mathbf{a}}_{\text{MAP}, \hat{\xi}_{\text{th}}, j}, \quad (45)$$

where $\hat{\mathbf{a}}_{\text{MAP}, \hat{\xi}_{\text{th}}, j}$ corresponds to the wavelet coefficients of the thresholded surrogate image $\hat{\mathbf{x}}_{\text{MAP}, \hat{\xi}_{\text{th}}}$ at level j . The errors at level j can be approximated by the difference between $\hat{\mathbf{x}}_{\text{MAP}}$ and $\hat{\mathbf{x}}_{\text{MAP}, \hat{\xi}_{\text{th}}, j}$.

There are two main advantages of this approach to pixel-based UQ with respect to the LCIs described in Section 4.2.2. The first one is the reduced computational complexity, as we only need to solve a single root-finding problem, significantly reducing the number of likelihood evaluations. The second is that when we saturate the HPD region, we consider the entire image simultaneously. In the LCI computation, we only change a small group of pixels until it saturates the HPD region that corresponds to the entire image. This behaviour can be problematic as, for example, the LCI error bounds will be larger if the size of the image grows and the superpixel size is kept the same, which is an undesirable property. Consequently, the predicted errors from the new pixel UQ approach are faster to compute and considerably tighter than the LCIs.

4.3 MCMC sampling and UQ validation

As stated before, MCMC sampling is not yet scalable to the high dimensions of the RI imaging problems we target. However, sampling is still helpful in validating the UQ approaches of this paper. If we sample from the posterior distribution, we can compute the posterior standard deviation, providing a visual representation of the posterior, including the learned convex regularizer. Sampling from the posterior also allows us to compare the MAP estimator

with another widely known estimator, the posterior mean (Arridge et al. 2019), which coincides with the minimum mean squared error (MMSE) estimator under some assumptions.

The log-posterior distribution of the QUANTIFAI model with the CRR-NN reads

$$-\log p_{\text{CRR-NN}}(\mathbf{x}|\mathbf{y}) \propto \frac{1}{2\sigma^2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \frac{\lambda}{\mu} R_{\theta}(\mu\mathbf{x}) + \ell_{\mathbb{R}^N}(\mathbf{x}), \quad (46)$$

with the first two terms belonging to \mathcal{C}^1 with Lipschitz gradients, we do not need to use any approximation, e.g. the MY envelope, to sample from it. The reality constraint is enforced directly when evaluating the gradient of the log-likelihood. The Langevin diffusion sampling algorithms reviewed in Section 2.5.3 require the gradient of the log-posterior distribution, which have been computed in equations (26) and (27). In practice, we will use the SK-ROCK algorithm (Pereyra et al. 2020) as it provides a faster convergence than the ULA algorithm. We omit the subsequent MH step to accelerate the calculations motivated by the consistent results from Cai et al. (2018a).

The log-posterior distribution of the analysis formulation of the model from Cai et al. (2018a) with a wavelet-based sparsity enforcing prior reads

$$-\log p_{\text{WAV}}(\mathbf{x}|\mathbf{y}) \propto \frac{1}{2\sigma^2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda \|\Psi^\dagger\mathbf{x}\|_1 + \ell_{\mathbb{R}^N}(\mathbf{x}), \quad (47)$$

which includes the non-smooth prior term with the ℓ_1 norm, and the reality constraint which we again apply to the gradient of the log-likelihood. We resort to the MY envelope γ -approximation of the sparsity-inducing prior term as shown in equation (14). The proximal operator of the prior term has a closed-form solution that reads

$$\mathcal{S}_{\text{soft}, \beta_{\text{th}}}(\mathbf{a}) = [\mathcal{S}_{\text{soft}, \beta_{\text{th}}}(a_1), \dots, \mathcal{S}_{\text{soft}, \beta_{\text{th}}}(a_L)]^T, \quad (48)$$

where $\mathbf{a} = \Psi^\dagger\mathbf{x}$ and we have applied element-wise the soft-thresholding function

$$\mathcal{S}_{\text{soft}, \beta_{\text{th}}}(a_i) = \begin{cases} 0, & \text{if } |a_i| \leq \beta_{\text{th}}, \\ \frac{a_i}{|a_i|} (|a_i| - \beta_{\text{th}}), & \text{otherwise.} \end{cases} \quad (49)$$

The threshold β_{th} used in practice is $\lambda\gamma$, the product of the regularization strength and the parameter of the MY approximation. See Cai et al. (2018a) for more details on sampling the model with a wavelet-based regularization. In practice, we again rely on the SK-ROCK algorithm for sampling and avoid using an MH step for the reasons mentioned above.

5 EXPERIMENTAL RESULTS

In this section, we demonstrate the QUANTIFAI model against the wavelet-based model presented in Cai et al. (2018a, b) as it is one of the few RI imaging methods providing UQ capabilities. We use a simulated setup with real reconstructed RI images considered as the ground truth. We focus on the UQ capabilities of the methods, while also considering reconstruction performance.

5.1 Data set

The base images used in our experiment are the ones from Cai et al. (2018a): (i) the HI region of the M31 galaxy in Fig. 1(a), (ii) the W28 supernova remnant in Fig. 2(a), (iii) the 3C288 radio galaxy in Fig. 3(a), and (iv) the Cygnus A radio galaxy in Fig. 4(a). All the images are 256×256 pixels, except for the Cygnus A galaxy, which is 256×512 . As the ground truth images are reconstructed from real observations, some minor original residuals and backgrounds

are more noticeable in the log scale images; for example, see Fig. 3(b). The ground truth images' values have been normalized to a unitless range between 0 and 1, and therefore, the colour bars in the reconstruction figures follow this range.

The previous images correspond to \mathbf{x} in our observational model from equation (3). The complex Gaussian noise $\mathbf{n} \in \mathbb{C}^M$ is composed of independent and identically distributed (*i.i.d.*) samples. Each sample is simulated following a complex Gaussian distribution, $n_i \sim \mathcal{N}_{\mathbb{C}}(0, \sigma^2)$, which implies that $\text{Re}(\mathbf{n}), \text{Im}(\mathbf{n}) \sim \mathcal{N}(0, \sigma/\sqrt{2})$ (Tse & Viswanath 2005). The noise is set such that we get a specific input signal-to-noise ratio (ISNR) on each image. For all the experiments, we set the ISNR to 30 dB, and the noise standard deviation is computed as follows

$$\sigma = \frac{\|\Phi\mathbf{x}\|_2}{\sqrt{M}} 10^{-\text{ISNR}/20}. \quad (50)$$

To mimic the uv plane coverage, we reuse the Fourier mask from Cai et al. (2018a, fig. 2) and use it to generate the visibilities from \mathbf{y} . The variable sampling density profile was taken from Puy, Vanderghenst & Wiaux (2011) and represents a 10 per cent coverage of the Fourier plane. In the experiments in the current section, we work with gridded visibilities where we have around 1.3×10^4 visibilities for Cygnus A and 6.5×10^3 visibilities for the rest of the images. The validation of the UQ techniques through MCMC sampling requires a large amount of iterations. The use of gridded visibilities allows us to base the forward operator Φ from equation (3) on the FFT (Cooley & Tukey 1965), helping us to alleviate the computational burden of the validation. Section 6 presents results with ungridded visibilities.

5.2 Models and experiment settings

5.2.1 RI imaging models

The CRR-NN in the QUANTIFAI model is a pre-trained model with $t = 5$, Gaussian white noise with standard deviation $\sigma = 5$, and parameters $\mu = 20$, $\lambda = 5 \times 10^4$. The model was trained on a set of natural images from the BSD500 data set (Arbeláez et al. 2011) containing images of landscapes, people, animals, and objects among others. The images are scaled to the $[0, 255]$ range, using ℓ_1 norm as the loss function in equation (22) with the Adam optimizer (Kingma & Ba 2017). The training parameters followed Goujon et al. (2023b, section VI.A).

The wavelet dictionary Ψ used in the wavelet-based model is composed of Daubechies 8 wavelets (Daubechies 1992) with a multiresolution level $J = 4$ following the setup from Cai et al. (2018a, b). The regularization parameter λ_{WAV} was set to 1×10^2 .

The regularization strengths of both models, λ and λ_{WAV} , were set to maximize the MAP reconstruction SNR using a grid search. We observed that QUANTIFAI's reconstruction SNR is significantly more robust with respect to the choice of regularization strength than the wavelet-based models.

5.2.2 Optimization settings

For QUANTIFAI, we used the optimization algorithm shown in Algorithm 1. The wavelet-based model also requires a proximal algorithm due to its non-smooth component and to provide a fair comparison we used the FISTA algorithm (Beck & Teboulle 2009) presented with more detail in Appendix A. In these experiments, we assumed that the noise level σ is known. If the noise level is unknown, it may be estimated by standard techniques (Price et al. 2021b). Both algorithms' tolerance criterion ξ was set to 10^{-5} , and the total number

of iterations to 1.5×10^4 . Nevertheless, both optimization algorithms converged before the total number of iterations was reached.

5.2.3 MCMC sampling settings

We generate 5×10^4 samples from each posterior distribution, with 5×10^4 burn-in iterations and a thinning factor of 10. The burn-in iterations consist of generating several samples that will be discarded so that the chain passes its transient period. The thinning factor corresponds to the number of samples that need to be generated between two samples so that they can be considered independently drawn from the target distribution. The sampling algorithm produced a total of 5.5×10^5 samples for each model. We have set to 10 the number of stages for the SK-ROCK algorithm (Pereyra et al. 2020), which is one of its main hyperparameters. The sampling of the posterior probability distributions is used as a validation, and therefore we set the sampling parameters focusing on good reconstructions and posterior samples rather than speed.

The wavelet-based model requires the MY envelope approximation to guarantee the chain's convergence, as described in Section 2.5.3 and Section 4.3. The MY approximation parameter γ was set to the inverse of the likelihood gradient's Lipschitz constant, c.f. the first term of equation (28).

The choice of the step sizes is critical to ensure the chains' convergence to the target distribution in a reasonable amount of time. The step size is chosen as a function of each posterior gradient's Lipschitz constant. The step sizes δ_Q and δ_W , corresponding to the QUANTIFAI and wavelet-based models, respectively, are computed as follows

$$\delta_Q = \frac{\kappa_Q}{L_{\text{likelihood}} + L_{\text{prior-CRR-NN}}}, \quad \delta_W = \frac{\kappa_W}{L_{\text{likelihood}} + \gamma^{-1}}, \quad (51)$$

where the Lipschitz constant bounds are shown in equation (28), and κ_Q and κ_W , are two positive constants smaller than one, here set to 0.98. We have followed the advice from Durmus et al. (2018) and Cai et al. (2018a) to set the sampling parameters.

5.2.4 UQ settings

We set $\alpha = 0.01$ in all the UQ methods, so the confidence level is 99 per cent. We used the bisection algorithm to compute the LCIs and the fast pixel UQ at different scales, with tolerance 10^{-4} and maximum number of iterations 200, for both models. We used the same wavelet dictionary as in the wavelet-based model for the fast pixel UQ at different scales.

The inpainting algorithm uses the same stopping criterion as Algorithm 1. In this case, the tolerance is set to 5×10^{-6} , and the total number of iterations to 1.5×10^4 . The CRR-NN used for the inpainting is the same one used in the QUANTIFAI model.

The Gaussian blurring kernel $G(0, \Sigma)$ from equation (35) is set using $\Sigma = \sigma_G^2 I_{2 \times 2}$, with σ_G being 3.5 pixels and a truncation radius of 7 pixels, giving a kernel $G \in \mathbb{R}^{15 \times 15}$.

5.3 Image reconstruction

We present the RI image reconstructions of the four ground truth test images in Figs 1, 2, 3, and 4. In each figure, we compare the wavelet-based and QUANTIFAI models, and we include the dirty reconstruction as a reference. The metric used to compare the RI image reconstruction is the SNR expressed in dB defined as follows

$$\text{SNR}(\mathbf{x}, \mathbf{x}_{\text{gt}}) = -20 \log_{10} \left(\frac{\|\mathbf{x}_{\text{gt}} - \mathbf{x}\|_2}{\|\mathbf{x}_{\text{gt}}\|_2} \right), \quad (52)$$

Table 1. Reconstruction performance of the different point estimates for the data set images in terms of SNR with respect to the ground truth. We compare the MAP and the MMSE reconstruction of the wavelet-based and the QUANTIFAI model. We include the dirty reconstruction as a reference. We observe that the MAP estimation from QUANTIFAI outperforms the other reconstructions from the wavelet-based prior and all the MMSE estimations.

Images	Reconstruction SNR (dB)				
	Dirty	Wavelet-based prior		QUANTIFAI	
		MMSE	MAP	MMSE	MAP
W28	3.39	18.17	23.04	23.38	26.85
M31	5.01	23.78	25.52	24.61	27.48
3C288	7.02	14.31	14.15	23.23	24.10
Cygnus A	4.60	20.52	17.53	25.36	30.25

where \mathbf{x}_{gt} corresponds to the reference or ground truth, and \mathbf{x} to the estimation, and $\|\cdot\|_2$ is the usual ℓ_2 norm.

The quantitative reconstruction performance results are presented in Table 1. The MAP reconstruction from QUANTIFAI performs significantly better than the wavelet-based counterpart in every image from our data set. The performance gain lies between 1.9 and 12.7 dB, with an average gain of 7 dB. It is difficult to see the QUANTIFAI improvements by eye when inspecting reconstructed images. However, when observing the errors in the fourth column, the improved quality of QUANTIFAI's reconstructions becomes evident. Shifting towards the sampling results, we observe a similar behaviour of the MMSE reconstruction in favour of QUANTIFAI's images. The MAP is considerably faster than the MMSE, relying on optimization rather than posterior sampling. Recall that the MMSE is built as averaging posterior samples. In addition, the MAP consistently provides improved reconstruction performance with respect to the MMSE.

The posterior standard deviation provides a qualitative way to validate the posterior model and its uncertainties. The comparison of the posterior standard deviation with the MAP reconstruction error shows a higher correlation for the QUANTIFAI model than the wavelet-based model. In addition, the posterior standard deviation of QUANTIFAI shows lower variance than its wavelet-based counterpart, which is in agreement with QUANTIFAI's smaller reconstruction error. For example, in image W28 in Fig. 2, we observe in sub-Fig. 2(j) that the posterior standard deviation value is large near the edges of the ground truth image. It is reassuring that QUANTIFAI's reconstruction error also shows the same behaviour.

The performance results showcase the expressive power of the CRR-NN-based prior even if the regularizer is constrained to be convex. The results also confirm the generalization power of the CRR-NN-based prior. Even if trained on natural images, the CRR-NN can provide remarkable reconstruction performances for astronomical images and meaningful posterior standard deviations.

The reconstructions using the wavelet-based prior model do exhibit some low-intensity artefacts in the Cygnus A and 3C288 images, as shown in Figs 4(h) and 3(h). These artefacts are due to the patterns in the ground truth images, which originate from real observations, and the thresholding of the orthogonal wavelet basis. Such patterns are absent in the M31 and W28 images because the noise was removed in a pre-processing step, as seen in Figs 1(b) and 2(b) in comparison to Figs 4(b) and 3(b). The regularization strength for the wavelet-prior model was selected to maximize the reconstruction SNR. This chosen value is lower than in Cai et al. (2018b), explaining the observed patterns and the finer details in our reconstructions. Employing a wavelet dictionary instead of an orthogonal wavelet basis and adding a positivity-enforcing constraint could mitigate the appearance of these artefacts.

5.4 Hypothesis testing of image structure

We start by carrying out hypothesis tests of image structure, which are the most scalable UQ techniques we will study. First, a surrogate image is created by modifying one region of interest. It only takes one further evaluation of the likelihood and prior potential to carry out the hypothesis test. The test helps us to quantitatively answer a scientific question with a $100(1 - \alpha)$ per cent confidence level. The scientific question targeted depends on the constructed surrogate image, and in this work, we consider two scenarios.

In the first scenario, we consider a particular structure in the reconstructed intensity image. We can query whether the structure's origin is physical or not. For example, the structure could be a reconstruction artefact or a physical process. Fig. 5 shows this option, where we have analysed different regions of the four images. The first four inpainted regions correspond to physical structures, and the fifth region, i.e. region number 2 of image 3C288, does not correspond to a physical structure. The surrogate images are produced with an inpainting algorithm using QUANTIFAI's prior so that the inpainted region agrees with the prior.

The second scenario is to blur the finer structure in the reconstructed image and perform a hypothesis test to elucidate the question of whether the blurred structure is physical or not. The test is illustrated in Fig. 6. In this case, all four blurred images represent physical structures.

In both cases, we compare the hypothesis test using a MAP-based approach described in this work and a sampling-based approach for validation. In the MAP-based approach, we build the HPD region in equation (31) with the approximation in equation (32) and use the MAP estimation as our reconstruction. In the sampling-based approach, we use the MMSE as the reconstruction, i.e. the mean of the posterior samples, and compute the threshold defining the HPD region using the quantile function on the potentials of the posterior samples following Cai et al. (2018a, Section 5.2).

Table 2 presents the results for the inpainting hypothesis test, where the inpainted surrogates are shown in Fig. 5. The MAP- and sampling-based results are consistent in all the images studied, where the threshold computed with the posterior samples is slightly tighter than the MAP-based approximation. The hypothesis tests correctly classify the structure in images M31, W28, and 3C288, including the two cases of the latter image. The UQ methods cannot make a strong statistical statement about the structures in the Cygnus A image. In this image, where the inpainted region has a tiny physical structure, the potentials of the inpainted surrogate image rest close to the MAP and MMSE estimators. We include the hypothesis test results of the same inpainting experiment for the wavelet-based model in Appendix B1 to provide a comparison between the models. We used the wavelet prior to inpaint the region of interest to allow for a fair comparison. All results from the wavelet-based model are in agreement with QUANTIFAI.

The results from the blurred surrogates of Fig. 6 are presented in Table 3. In all the images, the hypothesis test concludes that the blurred fine structure is physical as the potential falls out of the HPD region. The MAP- and sampling-based results are consistent with each other.

The different hypothesis tests have shown consistent results between the sampling-based and highly scalable MAP-based results. In addition, the results from the hypothesis tests are coherent between the QUANTIFAI and wavelet-based model. We remark that the approach based on the MAP requires one further measurement operator evaluation to carry out the hypothesis test. The test provides a highly scalable way to answer scientific questions about the uncertainty of the RI imaging reconstructions.

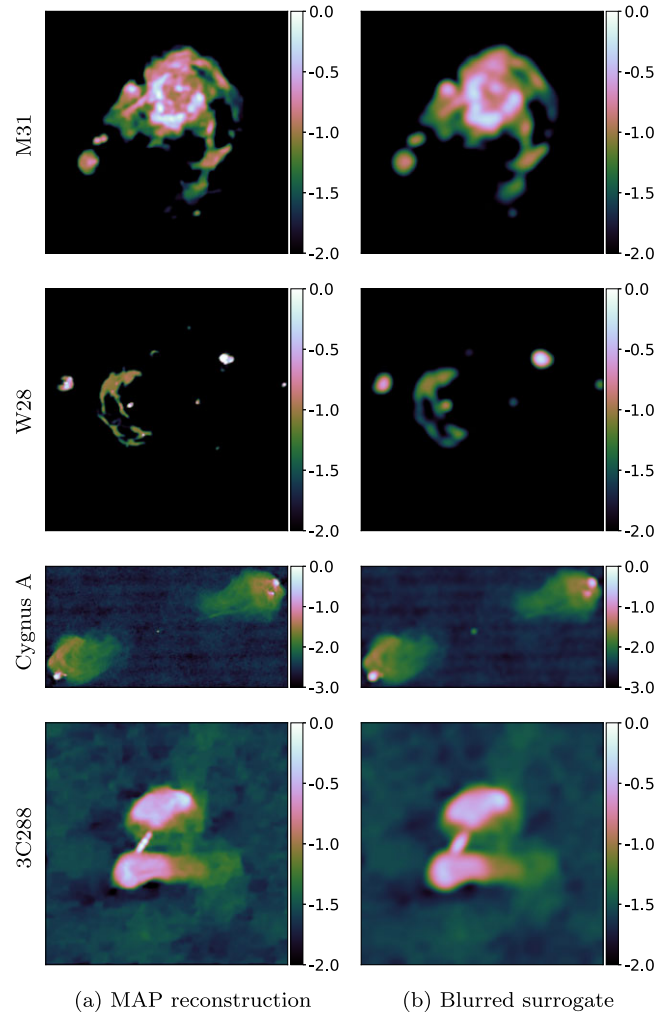


Figure 6. Hypothesis test of the fine structure in the four QUANTIFAI MAP reconstructions for M31, W28, Cygnus A, and 3C288. All the images are shown in \log_{10} scale. The fine structure is blurred using a Gaussian kernel with a standard deviation of 3.5 pixels and a radius of 7 pixels. The blurred surrogate images in the second column are constructed by blurring the MAP reconstruction shown in the first column. The hypothesis tests are done with the QUANTIFAI model.

5.5 Local credible intervals

We have exploited the approximation of the HPD region from Section 4.1 based on the MAP estimations and a credible level of 99 per cent. The approximate HPD regions were then used to compute the LCIs, whose lengths per pixel are visualized as an image, c.f. Fig. 7. The LCI lengths are displayed after subtracting the mean LCI length overall superpixels in the image, which is shown in the top left corner of the image. The UQ results for QUANTIFAI are presented for two superpixel sizes, 4×4 and 8×8 . We have omitted LCIs from the wavelet-based prior for conciseness. The posterior standard deviations at the two superpixel sizes are included for comparison with the significantly faster MAP-based UQ technique of the LCIs. We find a reasonable agreement between the structure in the LCI plots and the posterior standard deviation. For example, the 3C288 image with superpixel size 8×8 yields tighter LCIs in the two elliptical regions and in the small connecting structure in the centre of the image. The corresponding posterior standard deviation is smaller in the aforementioned regions, which is expected as most

Table 2. Hypothesis test results for the inpainted surrogates in Fig. 5 using the QUANTIFAI model. The function $(f + g)(\cdot)$ corresponds to the combined potential of the likelihood and the prior. The reconstruction \hat{x}^* represents the point estimate used in the sampling or optimization scenarios, which are the MMSE and the MAP, respectively. The SK-ROCK method corresponds to the posterior sampling techniques. The image $\hat{x}^{*,\text{sgt}}$ corresponds to the surrogate image, where the areas of interest shown in Fig. 5 have been inpainted. The isocontours, $\hat{\gamma}_{0.01}$, or thresholds, are calculated with an α of 0.01 giving a credible set of 99 per cent. In the MAP row, the threshold is computed following the approximation in equation (32). In the SK-ROCK row, the threshold is computed from the posterior samples following Cai et al. (2018a). The symbols \checkmark and \times in the Ground truth column indicate if the inpainted region contains a physical structure from the ground truth or not, respectively. In the last column, the \checkmark indicates that the hypothesis test is conclusive. All values are scaled with 10^5 . QUANTIFAI is able to correctly reject the hypothesis for all images where the tested structure is physical except for the Cygnus A image. In all cases, the MAP-based and MCMC sampling-based results agree with each other.

Images	Test area	Ground truth	Method	Point estimate $(f + g)(\hat{x}^*)$	Surrogate $(f + g)(\hat{x}^{*,\text{sgt}})$	Isocontour $\hat{\gamma}_{0.01}$	Hypothesis test
M31	1	\checkmark	SK-ROCK	0.340	1.159	0.742	\checkmark
			MAP	0.310	1.161	0.990	\checkmark
Cygnus A	1	\checkmark	SK-ROCK	0.125	0.182	0.848	\times
			MAP	0.105	0.169	1.450	\times
W28	1	\checkmark	SK-ROCK	0.222	4.649	0.612	\checkmark
			MAP	0.196	4.699	0.876	\checkmark
3C288	1	\checkmark	SK-ROCK	0.257	1.918	0.659	\checkmark
			MAP	0.229	1.908	0.908	\checkmark
	2	\times	SK-ROCK	0.257	0.257	0.659	\times
			MAP	0.229	0.229	0.908	\times

Table 3. Hypothesis test results for the blurred surrogates of Fig. 6 using the QUANTIFAI model. The description of Fig. 6 holds in this table. All values are scaled with 10^5 . QUANTIFAI is able to correctly reject the hypothesis in all cases, and the MAP-based outcome agrees with its MCMC sampling-based counterpart.

Images	Method	Initial $(f + g)(\hat{x}^*)$	Surrogate $(f + g)(\hat{x}^{*,\text{sgt}})$	Isocontour $\hat{\gamma}_{0.01}$	Hypothesis test
M31	SK-ROCK	0.340	1.905	0.742	\checkmark
	MAP	0.310	1.906	0.990	\checkmark
Cygnus A	SK-ROCK	0.125	9.642	0.848	\checkmark
	MAP	0.105	9.643	1.450	\checkmark
W28	SK-ROCK	0.222	8.389	0.612	\checkmark
	MAP	0.196	8.387	0.876	\checkmark
3C288	SK-ROCK	0.257	0.920	0.659	\checkmark
	MAP	0.229	0.922	0.908	\checkmark

of the observed signal concentrates there. The LCIs and the posterior standard deviation represent different quantiles, so we would not expect an exact agreement even without any approximation in the computation of the LCIs.

We observe, as expected, that the larger superpixels have tighter LCIs, as seen in the mean LCIs shown on the top left corner of the subfigures in Fig. 7. The reconstructions are naturally less uncertain on the larger scales due to the properties of our measurement operator, as the visibilities are generally concentrated towards the low frequencies. In addition, varying the value of a larger superpixel saturates the HPD region faster than for a small superpixel. We have also computed the LCIs for the superpixels of size 16×16 , which we have not included for conciseness. The corresponding mean LCI values are 0.20, 0.08, 0.24, and 0.07 for the images in the same order as in Fig. 7.

When comparing the mean value of the LCIs from the four reconstructions from Fig. 7 we notice that two of them, M31 and 3C288, have higher uncertainty than the rest. The higher the uncertainty, the larger the mean value of the LCI gets, as the superpixel values need larger changes before they saturate the HPD region. Image

3C288, with a superpixel size of 4×4 , is an example where the LCIs have saturated as the mean is close to unity;³ therefore, the LCI image's detailed structure is lost due to the saturation. This saturation highlights the need for superpixel sizes to be selected appropriately, depending on the case at hand.

5.6 Fast pixel uncertainty quantification at different scales

The fast pixel UQ method results for the images M31 and W28 are reported in Fig. 8. We use the error between the MAP estimation and the ground truth image, i.e. true error, to validate the predicted uncertainty of the fast UQ method. The true error at different scales can be computed following equation (45),

$$\hat{x}_{\text{GT},j} = \sum_{\substack{l=0, \\ l \neq j}}^J \Psi_l a_{\text{GT},l} + \Psi_j \hat{a}_{\text{MAP},j}, \quad (53)$$

³n.b. The images are scaled in the range [0,1].

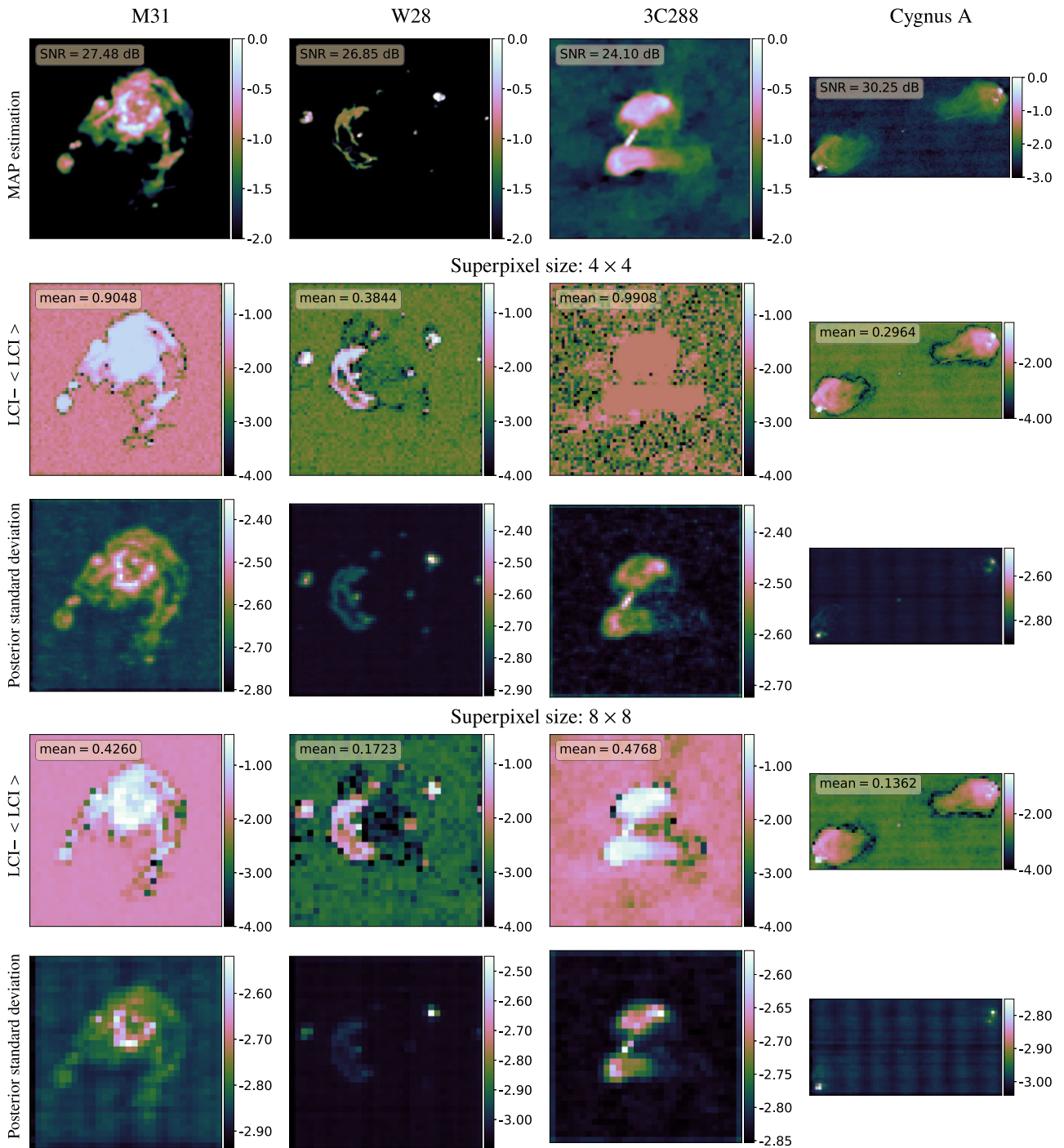


Figure 7. Length of the local credible intervals (LCIs), *c.f.* Bayesian error bars, computed with a 99 per cent credible level using superpixel sizes of 4×4 and 8×8 . Each column represents one of the four images in our data set. The first row shows the MAP estimation of each image at its original resolution. The second row displays the variation of the LCIs around their mean, recorded in a box in the upper left corners. This display choice allows us to visualize the structure of the LCIs better while keeping the LCIs mean information. The third row presents the posterior standard deviation computed with the same superpixel size. The fourth and fifth rows present the equivalent information for the superpixel size of 8×8 . There is reasonable agreement between the uncertainty captured by the LCI and the posterior standard deviation.

where $a_{GT,l}$ are the wavelet decomposition coefficients of the ground truth image at multi-resolution level l . We have replaced the ground truth image's wavelet coefficient at a single level with the coefficients from the MAP reconstruction.

We observe a good agreement between the predicted and ground truth errors at the different multi-resolution levels. There is an

overestimation of the errors, which can come from two sources. First, the approximation of the HPD region is conservative, as it has been discussed in Pereyra (2017). Secondly, the MAP estimation is already missing some of the fine or high-frequency structures in the ground truth images. This fact can be seen in the MAP reconstruction errors in subFigs 1(h) and 2(h). The missing high-frequency structure is

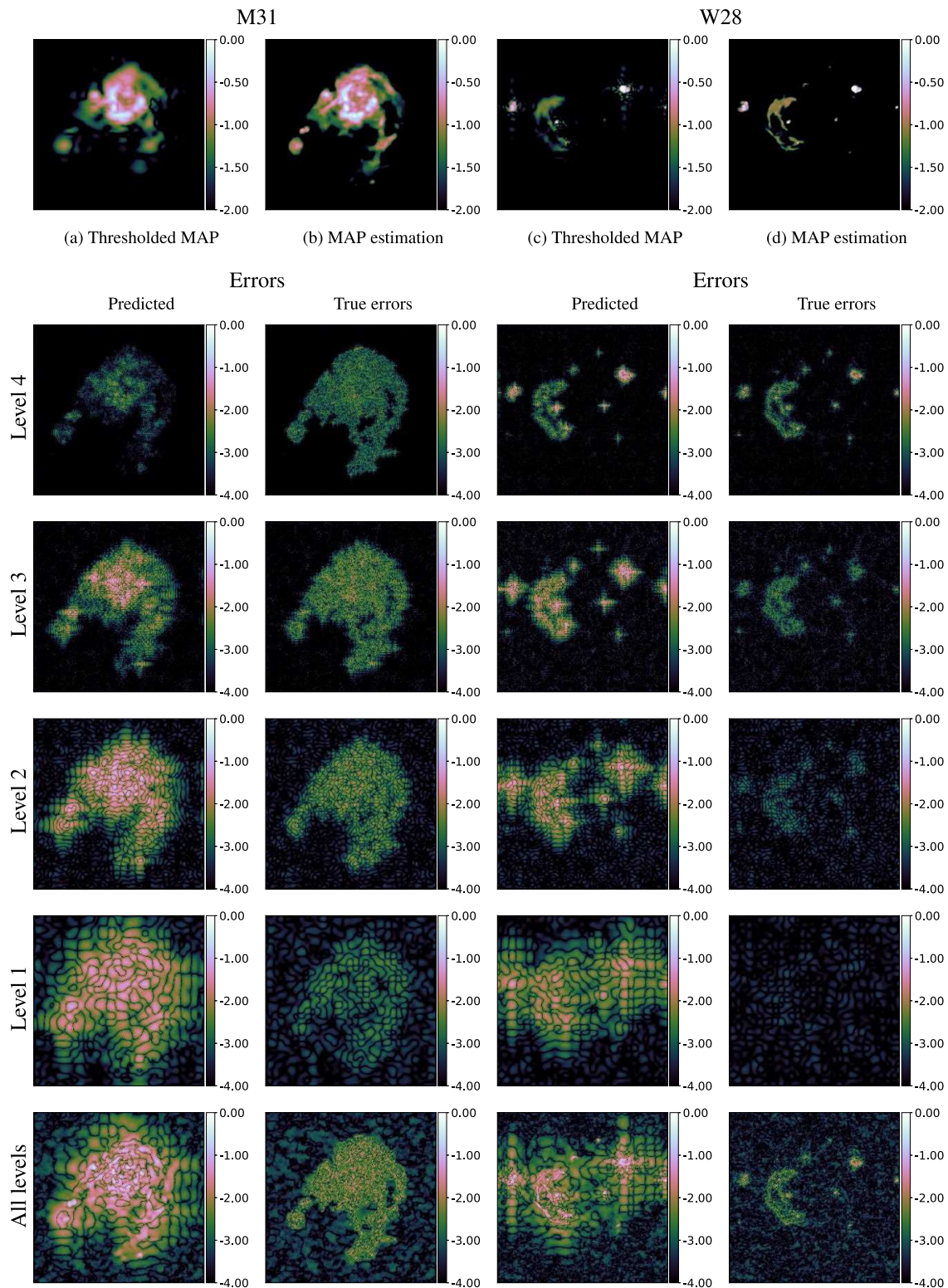


Figure 8. Fast pixel uncertainty quantification (UQ) with the QUANTIFAI model on the images M31 and W28. The first two columns correspond to M31, while the last two columns to W28. The first row displays the pairs of the thresholded MAP reconstruction that saturates the HPD region versus the original MAP reconstruction. The following rows compare the predicted error of the thresholded MAP computed with the fast pixel UQ method against the MAP reconstruction error using ground truth images at each wavelet scale. The last row shows the cumulative error when considering all scales.

Table 4. Computation wall-clock times for the W28 image in seconds for both models being compared.

Models	MAP optim.	Posterior sampling	LCIs 8×8	Fast pixel UQ
Wavelet-based	0.94	36.0×10^3	149.7	–
QUANTIFAI	0.64	6.44×10^3	108.2	0.17

Table 5. The number of measurement operator evaluations used by the QUANTIFAI for the W28 image. We do not distinguish between the measurement operator and its adjoint. Therefore, evaluating the log-likelihood gradient counts as two evaluations of the measurement operator. The fast pixel UQ is three and six orders of magnitude faster than the MCMC sampling and LCIs, respectively.

MCMC sampling	LCIs 8×8	LCIs 16×16	Fast pixel UQ
11×10^6	81.5×10^3	21.2×10^3	28

expected due to the properties of the measurement operator discussed in Section 2.1.

The structure from the chosen wavelet representation, Ψ , underpinning the UQ method can be observed in the predicted errors. This structure is visible mainly in the higher frequencies of the W28, where point sources are in the image. The wavelet structure should be taken into account when analysing the reconstruction errors.

This fast pixel UQ method allows us to approximate the reconstruction errors made at different scales for a fraction of the computational cost of the LCI pixel UQ method. The evaluations of the measurement operators are reduced by three orders of magnitude, resulting in an ultra-fast and truly scalable pixel UQ method. Furthermore, a single non-linear equation solve, e.g. root finding problem, of the new pixel UQ method suffices to predict the errors at all scales, while with LCIs, we are required to repeat the process for each superpixel size.

5.7 Computation time

The computation wall-clock time for both models, QUANTIFAI and the wavelet-based, are summarized in Table 4. We include the results for only the W28 image in Tables 4 and 5 as they are representative of the other images. All the computations for both models were done using an Nvidia-A100 40GB GPU using PYTORCH (Paszke et al. 2019). We observe a lower computation time for the QUANTIFAI model with respect to its wavelet-based counterpart. One reason is the lightweight CRR-NN model that quickly evaluates its gradient and potential. Note that the regularization strength has an impact on the number of iterations and it could be changed to favour a faster convergence. The regularization strength was chosen to optimize MAP reconstruction quality.

The results shown in Table 4 highlight the importance of relying on optimization-based rather than sampling-based reconstructions when focusing on the scalability of the method. There is a difference of approximately four orders of magnitude in the computation time of the MAP and the MMSE which relies on MCMC sampling techniques. Focusing on UQ, the posterior sampling is 60 times slower than the computation of the LCIs with 8×8 superpixels and more than 37 500 times slower than the fast pixel UQ proposed in this work. The new fast pixel UQ provides an extremely rapid approach to providing pixel-based UQ, over 630 times faster than the 8×8 LCIs.

The evaluation of the measurement operator is the most time-consuming operation in a real large-scale RI imaging scenario. If we target scalability, we need to monitor the number of measurement operator evaluations. Table 5 summarizes the number of measurement operator evaluations required for the UQ techniques. The results are only shown for the QUANTIFAI model as they are representative of the wavelet-based model. As mentioned before, we note the reduction of evaluations between optimization and sampling-based reconstructions. We remark on the reduction in the number of evaluations for the UQ tasks, approximately three orders of magnitude between the sampling and the LCIs, and three subsequent orders of magnitude between the LCIs and the fast pixel UQ. These results make the fast pixel UQ six orders of magnitude faster than MCMC sampling. The MAP estimation for the CRR required 1082 measurement operator evaluations. However, the algorithm's settings were chosen to maximize the reconstruction SNR. By modifying the regularization parameter of the CRR-based prior, we can reduce the number of evaluations by an order of magnitude. Recent developments in code parallelization for RI imaging reconstruction algorithms⁴ (Pratley et al. 2019a; Pratley & McEwen 2019) could be integrated to push the scalability of the method further.

6 EXPERIMENTAL RESULTS WITH UNGRIDDED VISIBILITIES

In the previous section, we have validated the proposed UQ methods from QUANTIFAI in a simple setting with gridded visibilities. This choice lets us use the FFT algorithm for the forward model, which allows us to run MCMC algorithms for UQ validation in a sensible amount of time. In this section, we showcase QUANTIFAI in an experiment using simulated visibility patterns from the MeerKAT radio telescope (Jonas & MeerKAT Team 2016). The main difference is that the visibility patterns are ungridded following a realistic distribution, which obliges us to rely on the Non-Uniform FFT (NUFFT) for the forward model. We remark that in this experiment, we are not addressing many of the challenges of dealing with real data, which are beyond the scope of this article.

6.1 Data set and experiment settings

We have simulated four single-frequency ungridded visibility patterns of differing synthesis times for MeerKAT. The start frequency is set to 1400 MHz with a channel width of 10 MHz. The pointing position has been randomly selected and set to (13h18m54.86s, $-15d36m04.25s$) in the J2000 reference, and it was maintained for the four generated data sets. We have used a publicly available code⁵ that is based on the CASA simulation software (The CASA Team 2022). The synthesis times used are 1, 2, 4, and 8 h, with a constant integration time of 240 s. Each data set has a field of view of approximately 1 deg^2 . The number of visibilities of each data set is 3×10^4 , 6×10^4 , 1.2×10^5 , and 2.4×10^5 , correspondingly. Fig. 9 presents the simulated visibility patterns. We reuse the images described in Section 5.1 as the ground truth, keeping their original dimensions, which are unrealistically small to be representative of a MeerKAT observation grid size.

To cope with the ungridded visibilities, we have to change the forward operator Φ from equation (3) that before was based on the FFT. We rely on the PYTORCH-based NUFFT implementation from

⁴<https://github.com/astro-informatics/purify> and <https://github.com/astro-informatics/sopt>

⁵<https://github.com/ratt-ru/simms>

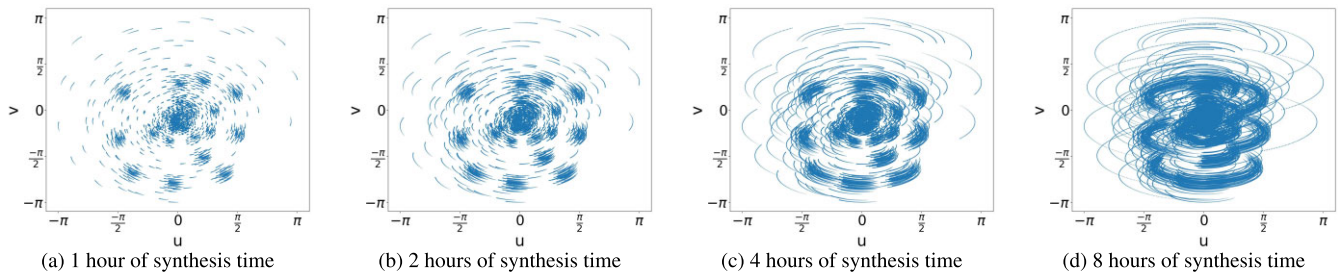


Figure 9. The four sets of simulated ungridded visibilities for the MeerKAT radio telescope with synthesis times of 1, 2, 4, and 8 h.

Muckley et al. (2020)⁶ based on Kaiser-Bessel gridding (Fessler & Sutton 2003). We have used the same images, Gaussian noise model presented in Section 5.1, and trained CRR-NN in QUANTIFAI. We have reused the previously introduced hyperparameter values of QUANTIFAI except for λ , which we have tuned to maximize the reconstruction's SNR for the four data sets to 10^4 , 1.4×10^4 , 1.9×10^4 , and 2.25×10^4 , respectively.

6.2 Results

The reconstructions and the fast pixel UQ maps for the image M31 are shown in Fig. 10. Each column corresponds to each of the four data sets. The results for the other images are postponed to Appendix C. Table 6 presents quantitative results regarding reconstruction SNR, measurement operator evaluations, and wall-clock computing time. The reconstructions present good quality in terms of SNR, which increases with longer synthesis times, which was expected as the Fourier coverage increases, as seen in Fig. 9. The wall-clock reconstruction time increases with the longer synthesis times as the time for each evaluation of the measurement operator becomes longer due to the larger number of visibilities.

Even in this more challenging setting, we find a good correlation between the predicted and the oracle error. The pixel UQ maps still show some patterns characteristic of the multiscale wavelet representation used and should, therefore, be considered in the analysis of the pixel UQ maps. The fast pixel UQ represents a fraction of the time and number of measurement operator evaluations required for the MAP reconstruction.

7 DISCUSSION

In this work, we have worked with synthetic data sets, starting with gridded visibilities and ending with realistic ungridded visibility patterns. The handling of real or realistic data is beyond the scope of this work, which focuses on the methodology presented and the validation of the UQ techniques. Some problems faced while handling large-scale observation include the calibration of direction-independent and direction-dependent effects. Further studies of the performance of QUANTIFAI to more realistic images with differing large dynamic ranges and bright point sources are left for the future. The extension of QUANTIFAI to incorporate a frequency axis in the reconstruction to cope with multi-frequency observations is also left for further study. Even if QUANTIFAI could reconstruct images with UQ maps with $\sim 10^5$ visibilities in less than two minutes, there is ongoing work to exploit existing C++ parallelization capabilities to scale the method further. A detailed performance and computing time benchmark is expected for the future implementation.

⁶<https://github.com/mmuckley/torchkbnufft>

The current approach to set the regularization strength of the CRR-NN, λ , with a grid search is not compatible with real data as we require access to the ground truth. There are several ways to circumvent this problem in the large-scale setting. One way forward is to consider a subset of the observations to alleviate the computational burden and rely on the empirical Bayesian approach from Vidal et al. (2020) and De Bortoli et al. (2020) to estimate the regularization parameter directly from the observed data. Another way forward is to follow a heuristic approach similar to Terris et al. (2022). The study of the best strategy for QUANTIFAI is left for further study.

We have not included a positivity constraint in the QUANTIFAI model, as is the case for Cai et al. (2018b), which we use for comparison. However, there is no impediment to adding a positivity constraint to QUANTIFAI; it amounts to changing the indicator function from equation (25) to an indicator function of the real positive orthant. The proximal operator associated with this indicator function has a closed form: the projection to the positive orthant. Consequently, we can compute the corresponding MAP solution and evaluate the potential using the modified indicator function.

The fast pixel UQ proposed in this article improved the tightness of the UQ bounds with respect to the LCIs by considering the entire image when saturating the HPD region. Nevertheless, these pixel-level UQ maps are intended to be shared as visual aids accompanying the reconstructions. The pixel UQ maps can drive the astronomer's attention to a specific region in the image where a consequent hypothesis test can be carried out using the techniques described in this work. The choice of wavelet basis impacts the fast pixel UQ maps produced, so it should be considered when analysing the maps. A way to improve the fast pixel UQ maps would be to replace the orthogonal wavelet basis used with a more performant dictionary as in SARA (Carrillo et al. 2012), which has shown to be well adapted for RI images.

8 CONCLUSIONS

In this work, we propose a new method coined QUANTIFAI that addresses uncertainty quantification in radio-interferometric (RI) imaging with data-driven (learned) priors in very high-dimensional settings. We have focused on three fundamental points in the RI imaging pipeline: scalability, estimation performance, and uncertainty quantification (UQ).

Our model builds upon a principled Bayesian framework for the UQ analysis, which is known to be computationally expensive when exploiting MCMC sampling methods. However, in this work, we leverage convex optimization techniques to estimate the MAP, the point estimate of the posterior distribution we use as reconstruction. We restrict our model to a log-concave posterior distribution to remain highly scalable and have Bayesian UQ techniques. This restriction is equivalent to having convex potentials for our likelihood

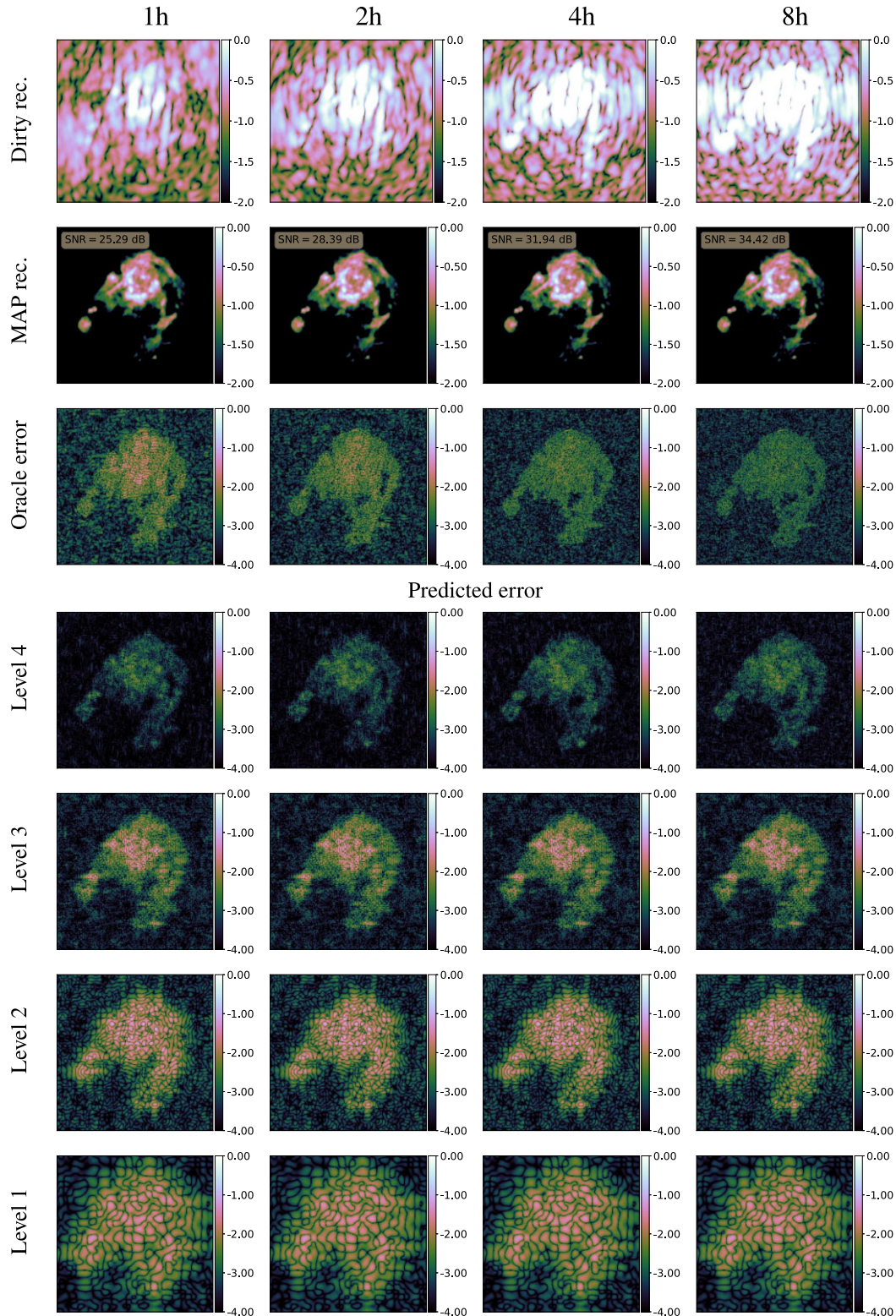


Figure 10. Reconstructions and fast pixel uncertainty quantification (UQ) with the QUANTIFAI model for the M31 image with the four sets of simulated MeerKAT ungridded visibilities. Each column corresponds to the four data sets with synthesis times of 1, 2, 4, and 8 h for a field of view of approximately 1 deg^2 . The first row represents the dirty reconstruction. The MAP reconstruction is presented in the second row, while the oracle error, which we do not have access to with real data, is shown in the third row. The different decomposition levels of pixel UQ are shown in the last four rows.

Table 6. Main results of QUANTIFAI for the M31 image with the realistic ungridded MeerKAT visibility patterns with differing synthesis times. As the number of visibilities grows with the synthesis timer, so does the reconstruction SNR. The number of visibilities increases proportionally to the synthesis times.

Metrics		Data sets			
		1 h	2 h	4 h	8 h
Reconstruction	Number of visibilities	3×10^4	6×10^4	1.2×10^5	2.4×10^5
	MAP reconstruction SNR (dB)	25.29	28.39	31.94	34.42
	Measurement op. evaluations	3288	2916	3006	3114
	Wall-clock time (s)	17.01	28.19	53.24	105.25
UQ	Measurement op. evaluations	26	28	30	30
	Wall-clock time (s)	0.28	0.44	0.73	1.26

and prior. In this scenario, we can exploit an approximation of the HPD region, which only requires the MAP estimation (Pereyra 2017) and bypasses expensive sampling techniques.

We want to include data-driven priors that can encode complex information learned implicitly from training data making them more expressive. Consequently, the learned priors allow us to improve performance with respect to previous models based on handcrafted priors (Cai et al. 2018b), e.g. wavelet-based sparsity-promoting priors. To support fast UQ techniques, our models must be convex, hence we adopt the recently introduced learnable convex-ridge regularizer neural network (CRR-NN; Goujon et al. 2023b). The CRR-NN-based prior is performant, reliable, and has shown to be robust to data distribution shifts. The QUANTIFAI model uses an analytic physically motivated model for the likelihood and the learned CRR-NN-based prior. In this work, we are focusing on the methodology, which is why we have only considered small problems, i.e. images of 256×256 . Nevertheless, QUANTIFAI can be integrated into the distributed frameworks (Pratley et al. 2019a; Pratley & McEwen 2019), which is the focus of ongoing work.

Numerical experiments are conducted with four images representative of RI imaging. We compare the QUANTIFAI model with the model containing a wavelet-based prior of Cai et al. (2018b). Our results show a considerable improvement in the reconstruction performance for QUANTIFAI. We validate our results against posterior samples from MCMC sampling algorithms and compute the posterior standard deviation. We found that QUANTIFAI produced more meaningful posterior standard deviations in comparison to the wavelet-based model. We also included numerical experiments with simulated MeerKAT ungridded visibilities, where we present QUANTIFAI's performance and computing times going up to $\sim 10^5$ visibilities.

We explore several MAP-based UQ techniques that rely on the approximate HPD region. We carry out hypothesis tests of image structure to assess if some structures observed in the reconstructions are physical. We then computed LCIs (c.f. Bayesian error bars) to measure the pixel-wise uncertainty. These two approaches were proposed by Cai et al. (2018b), and in this work, we validated them with MCMC posterior sampling results. Even if LCIs represent an already scalable alternative to sampling-based methods to provide pixel-wise UQ, they remain expensive for SKA-size data. Therefore, we proposed a novel pixel-wise UQ technique to approximate pixel errors at different scales that is three orders of magnitude faster than the LCIs. The new approach is based on thresholding the coefficients of a wavelet representation of the reconstruction until the HPD region saturates and is six orders of magnitude faster than sampling-based techniques.

QUANTIFAI proposed an approach with the potential to be highly scalable and performant to address UQ in RI imaging. In this work,

we have compared QUANTIFAI to a wavelet-based model using numerical experiments and a variety of metrics. However, as both models rely on the Bayesian framework, we could make a Bayesian model comparison, a principled approach to model selection and determine which model the data favours. Recent developments in McEwen et al. (2023) extend the model comparison to the learnt setting, with data-driven priors. The focus of ongoing work is to implement the proposed methodology in existing RI imaging frameworks PURIFY⁷ (Carrillo et al. 2014; Pratley et al. 2018, 2019b) and SOPT⁸ (Carrillo et al. 2012; Onose et al. 2016) to exploit massively parallelized computing environment (Pratley et al. 2019a; Pratley & McEwen 2019) and to realize the potential of scalability. In the near future, we plan to benchmark the speed and scalability of QUANTIFAI in a highly realistic setting.

A new perspective is to relax the convexity constraint of the prior by exploiting the fact that the posterior potential needs to be convex (rather than the prior) and that the RI imaging likelihood is already strongly convex. The relaxation of the CRR regularizer has been studied in a very recent work (Goujon, Neumayer & Unser 2023a), where a weakly-convex-ridge-regularizer neural network (WCRR-NN) has been proposed. If the WCRR-NN is adopted, it could further enhance the expressiveness of the regularizer and the reconstruction performance of QUANTIFAI in the RI imaging problem.

ACKNOWLEDGEMENTS

This work is supported by the UK Research and Innovation (UKRI) and Engineering and Physical Sciences Research Council (EPSRC) grant numbers EP/W007673/1 (LEXCI) and EP/T007346/1 (BOLT). MM is supported by the UCL Centre for Doctoral Training in Data Intensive Science (STFC Training grant ST/P006736/1).

We acknowledge the Python packages used in this work: IPYTHON (Pérez & Granger 2007), JUPYTER (Kluyver et al. 2016), MATPLOTLIB (Hunter 2007), PYTORCH (Paszke et al. 2019), NUMPY (Harris et al. 2020), ASTROPY (Astropy Collaboration 2022), SCIKIT IMAGE (van der Walt et al. 2014).

DATA AVAILABILITY

We provide the QUANTIFAI-PYTORCH-based Python package in a publicly available repository.⁹ We are in favour of reproducible research, which is why the images, visibilities, scripts, trained CRR-NN model and code to reproduce this article's experiments can be found in the aforementioned repository.

⁷<https://github.com/astro-informatics/purify>

⁸<https://github.com/astro-informatics/sopt>

⁹<https://github.com/astro-informatics/QuantifAI>

REFERENCES

- Abdulle A., Almuslimani I., Vilmart G., 2018, *SIAM/ASA J. Uncertain. Quan.*, 6, 937
- Adam A., Coogan A., Malkin N., Legin R., Perreault-Levasseur L., Hezaveh Y., Bengio Y., 2022, in *Machine Learning and the Physical Sciences Workshop, 36th conference on Neural Information Processing Systems (NeurIPS 2022)*, p. E1
- Adler J., Öktem O., 2018, *IEEE Trans. Med. Imaging*, 37, 1322
- Aghabiglou A., Terris M., Jackson A., Wiaux Y., 2023, in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, p. 1
- Aghabiglou A., San Chu C., Dabbech A., Wiaux Y., 2024, *ApJS*, 273, 3
- Allam T., 2016, Master's thesis, University College London
- Amos B., Xu L., Zico Kolter J., 2017, *Proc. 34th International Conference on Machine Learning*, Vol. 70. PMLR, p. 146
- Arbeláez P., Maire M., Fowlkes C., Malik J., 2011, *IEEE T. Pattern Anal. Mach. Intell.*, 33, 898
- Arras P., Knollmüller J., Junklewitz H., Enßlin T. A., 2018, 26th European Signal Processing Conference (EUSIPCO), Rome, Italy, p. 2683
- Arras P., Bester H. L., Perley R. A., Leike R., Smirnov O., Westermann R., Enßlin T. A., 2021, *A&A*, 646, A84
- Arras P., Frank P., Haim P., Knollmüller J., Leike R., Reinecke M., Enßlin T., 2022, *Nat. Astron.*, 6, 259
- Arridge S., Maass P., Öktem O., Schönlieb C.-B., 2019, *Acta Numer.*, 28, 1
- Ashton G. et al., 2022, *Nat. Rev. Methods Primers*, 2, 39
- Astropy Collaboration, 2022, *ApJ*, 935, 167
- Beck A., Teboulle M., 2009, *SIAM J. Imaging Sci.*, 2, 183
- Bhatnagar S., Cornwell T. J., 2004, *A&A*, 426, 747
- Bohra P., Campos J., Gupta H., Aziznejad S., Unser M., 2020, *IEEE Open J. Signal Process.*, 1, 295
- Boyd S., Parikh N., Chu E., Peleato B., Eckstein J., 2011, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*
- Braun R., Bourke T., Green J. A., Keane E., Wagg J., 2015, *Proc. Sci., Advancing Astrophysics with the Square Kilometre Array*. SISSA, Trieste, PoS#174
- Burden R. L., Faires J. D., 1989, *Numerical Analysis*, 4th edn. The Prindle, Weber and Schmidt Series in Mathematics. PWS-Kent Publishing Company, Boston
- Cai X., Pereyra M., McEwen J. D., 2018a, *MNRAS*, 480, 4154
- Cai X., Pereyra M., McEwen J. D., 2018b, *MNRAS*, 480, 4170
- Cai X., McEwen J. D., Pereyra M., 2022, *Stat. Comput.*, 32, 87
- Carrillo R. E., McEwen J. D., Wiaux Y., 2012, *MNRAS*, 426, 1223
- Carrillo R. E., McEwen J. D., Wiaux Y., 2014, *MNRAS*, 439, 3591
- Chambolle A., Pock T., 2011, *J. Math. Imaging Vision*, 40, 120
- Combettes P. L., Pesquet J.-C., 2011, in Bauschke H., Burachik R., Combettes P., Elser V., Luke D., Wolkowicz H., eds, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer Optimization and Its Applications. Vol 49, Springer, New York
- Condat L., 2013, *J. Optimiz. Theory Appl.*, 158, 460
- Connor L., Bouman K. L., Ravi V., Hallinan G., 2022, *MNRAS*, 514, 2614
- Cooley J., Tukey J. W., 1965, *Math. Comput.*, 19, 297
- Cornwell T. J., 2008, *IEEE J. Sel. Top. Signal Process.*, 2, 793
- Dabbech A., Ferrari C. Mary D., Slezak E. Smirnov O., Kenyon J. S., 2015, *A&A*, 576, A7
- Dabbech A., Onose A., Abdulaziz A., Perley R. A., Smirnov O. M., Wiaux Y., 2018, *MNRAS*, 476, 2853
- Daubechies I., 1992, *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia
- Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, *Proc. IEEE*, 97, 1482
- De Bortoli V., Durmus A., Pereyra M., Vidal A. F., 2020, *SIAM J. Imaging Sci.*, 13, 1990
- Dia N., Yantovski-Barth M. J., Adam A., Bowles M., Lemos P., Scaife A. M. M., Hezaveh Y., Perreault-Levasseur L., 2023, preprint (arXiv:2311.18012)
- Dong C., Loy C. C., He K., Tang X., 2016, *IEEE Trans. Pattern Anal. Mach. Intell.*, 38, 295
- Durmus A., Moulines E., Pereyra M., 2018, *SIAM J. Imaging Sci.*, 11, 473
- Durmus A., Moulines E., Pereyra M., 2022, *SIAM Rev.*, 64, 991
- Event Horizon Telescope Collaboration, 2019a, *ApJ*, 875, L1
- Event Horizon Telescope Collaboration, 2019b, *ApJ*, 875, L4
- Fessler J., Sutton B., 2003, *IEEE Trans. Signal Process.*, 51, 560
- Gheller C., Vazza F., 2021, *MNRAS*, 509, 990
- Golub G. H., van Loan C. F., 2013, *Matrix Computations*, 4th edn. JHU Press, Philadelphia, PA, available at: <http://www.cs.cornell.edu/cv/GVL4/golubandvanloan.htm>
- Goujon A., Neumayer S., Unser M., 2023a, *SIAM J. Imaging Sci.*, 17, 91
- Goujon A., Neumayer S., Bohra P., Ducotterd S., Unser M., 2023b, *IEEE Trans. Comput. Imaging*, 9, 781
- Hallinan G. et al., 2019, *BAAS*, 51, 255
- Harris C. R. et al., 2020, *Nature*, 585, 357
- Hendrycks D. et al., 2020, preprint (arXiv:2006.16241)
- Ho J., Jain A., Abbeel P., 2020, in Larochelle H., Ranzato M., Hadsell R., Balcan M. F., Lin H., eds, *Advances in Neural Information Processing Systems*, Vol 33. Curran Associates, Inc., p. 6840
- Högbom J. A., 1974, *A&AS*, 15, 417
- Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90
- Jonas J., Team MeerKAT, 2016, in *Proc. Sci., MeerKAT Science: On the Pathway to the SKA*. SISSA, Trieste, PoS#1
- Junklewitz H., Bell M. R., Selig M., Enßlin T. A., 2016, *A&A*, 586, A76
- Kingma D. P., Ba J., 2017, preprint (arXiv:1412.6980)
- Klatzer T., Dobson P., Altmann Y., Pereyra M., María Sanz-Serna J., Zygalakis K. C., 2024, *SIAM J. Imag. Sci.*, 17, 1078
- Kloeden P., Platen E., 2011, *Numerical Solution of Stochastic Differential Equations*. Stochastic Modelling and Applied Probability. Springer, Berlin
- Kluyver T. et al., 2016, in Loizides F., Schmidt B., eds, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press, Amsterdam, p. 87
- Knollmüller J., Enßlin T. A., 2019, preprint (arXiv:1901.11033)
- Knollmüller J., Arras P., Enßlin T., 2023, preprint (arXiv:2310.16889)
- Lunz S., Öktem O., Schönlieb C.-B., 2018, in Bengio S., Wallach H., Larochelle H., Grauman K., Cesa-Bianchi N., eds, *Advances in Neural Information Processing Systems*, Vol. 31. Curran Associates, Inc.
- Mallat S., 2008, *A Wavelet Tour of Signal Processing*, 3rd edn. The Sparse Way, 3rd edn. Academic Press, Inc., USA
- Mars M., Betcke M. M., McEwen J. D., 2023, *RAS Techn. Instrum.*, 2, 760
- Mars M., Betcke M. M., McEwen J. D., 2024, preprint (arXiv:2405.08958)
- McEwen J. D., Wiaux Y., 2011, *MNRAS*, 413, 1318
- McEwen J. D., Wallis C. G. R., Price M. A., Docherty M. M., 2021, preprint (arXiv:2111.12720)
- McEwen J. D., Liaudat T. I., Price M. A., Cai X., Pereyra M., 2023, *Phys. Sci. Forum*, 9, 13
- Melidonis S., Dobson P., Altmann Y., Pereyra M., Zygalakis K., 2023, *SIAM J. Imaging Sci.*, 16, 1195
- Monga V., Li Y., Eldar Y. C., 2021, *IEEE Signal Processing Magazine*, 38, 18
- Muckley M. J., Stern R., Murrell T., Knoll F., 2020, in *ISMRM Workshop on Data Sampling & Image Reconstruction*
- Mukherjee S., Dittmer S., Shumaylov Z., Lunz S., Öktem O., Schönlieb C.-B., 2020, preprint (arXiv:2008.02839)
- Nesterov Y., 2018, *Lectures on Convex Optimization*, 2nd edn. Springer Publishing Company, Cham
- Offringa A. R., Smirnov O., 2017, *MNRAS*, 471, 301
- Offringa A. R. et al., 2014, *MNRAS*, 444, 606
- Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J.-P., Pesquet J.-C., Wiaux Y., 2016, *MNRAS*, 462, 4314
- Parikh N., Boyd S., 2014, *Found. Trends Optim.*, 1, 127
- Paszke A. et al., 2019, preprint (arXiv:1912.01703)
- Pereyra M., 2016, *Stat. Comput.*, 26, 745

Pereyra M., 2017, *SIAM J. Imaging Sci.*, 10, 285
Pereyra M., 2019, *SIAM J. Imaging Sci.*, 12, 650
Pereyra M., Miele L. V., Zygalakis K. C., 2020, *SIAM J. Imaging Sci.*, 13, 905
Pereyra M., Vargas-Miele L. A., Zygalakis K. C., 2023, *SIAM J. Imag. Sci.*, 16, 2040
Pérez F., Granger B. E., 2007, *Comput. Sci. Eng.*, 9, 21
Perley R. A., Chandler C. J., Butler B. J., Wrobel J. M., 2011, *ApJ*, 739, L1
Pesquet J.-C., Repetti A., Terris M., Wiaux Y., 2021, *SIAM J. Imaging Sci.*, 14, 1206
Polanska A., Price M. A., Spurio Mancini A., McEwen J. D., 2023, *Phys. Sci. Forum*, 9, 10
Pratley L., McEwen J. D., 2019, preprint (arXiv:1903.07621)
Pratley L., McEwen J. D., d’Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2018, *MNRAS*, 473, 1038
Pratley L., McEwen J. D., d’Avezac M., Cai X., Perez-Suarez D., Christidi L., Guichard R., 2019a, preprint (arXiv:1903.04502)
Pratley L., Johnston-Hollitt M., McEwen J. D., 2019b, *ApJ*, 874, 174
Price M. A., Cai X., McEwen J. D., Pereyra M., Kitching T. D., (for the LSST Dark Energy Science Collaboration), 2019, *MNRAS*, 492, 394
Price M. A., Pratley L., McEwen J. D., 2021a, preprint (arXiv:2105.04935)
Price M. A., McEwen J. D., Cai X., Kitching T. D., Wallis C. G. R., (for the LSST Dark Energy Science Collaboration), 2021b, *MNRAS*, 506, 3678
Puy G., Vanderghyest P., Wiaux Y., 2011, *IEEE Signal Process. Lett.*, 18, 595
Rau U., Bhatnagar S., Voronkov M. A., Cornwell T. J., 2009, *Proc. IEEE*, 97, 1472
Remy B., Lanusse F., Jeffrey N. Liu J., Starck J.-L., Osato K., Schrabback T., 2023, *A&A*, 672, A51
Repetti A., Pereyra M., Wiaux Y., 2019, *SIAM J. Imaging Sci.*, 12, 87
Robert C., 2007, *The Bayesian Choice: From Decision Theoretic Foundations to Computational Implementation*. Springer, New York
Ronneberger O., Fischer P., Brox T., 2015, preprint (arXiv:1505.04597)
Roth J., Arras P., Reinecke M., Perley R. A., Westermann R., Enßlin T. A., 2023, *A&A*, 678, A177
Ryu E., Liu J., Wang S., Chen X., Wang Z., Yin W., 2019, in Chaudhuri K., Salakhutdinov R., eds, *Proceedings of Machine Learning Research Vol. 97, Proceedings of the 36th International Conference on Machine Learning*. PMLR, Long Beach, CA, p. 5546. Available at: <https://proceedings.mlr.press/v97/ryu19a.html>
Schmidt K., Geyer F., Fröse S., Blomenkamp P. S., Brüggem M., de Gasperin F., Elsässer D., Rhode W., 2022, *A&A*, 664, A134
Skilling J., 2006, *Bayesian Anal.*, 1, 833
Smirnov O. M., 2011a, *A&A*, 527, A106
Smirnov O. M., 2011b, *A&A*, 527, A107
Smirnov O. M., 2011c, *A&A*, 527, A108
Smirnov O. M., 2011d, *A&A*, 531, A159
Song Y., Sohl-Dickstein J., Kingma D. P., Kumar A., Ermon S., Poole B., 2020, preprint (arXiv:2011.13456)
Spurio Mancini A., Docherty M. M., Price M. A., McEwen J. D., 2023, *RASTI*, 2, 710
Starck J.-L., Murtagh F., Fadili J. M., 2010, *Sparse Image and Signal Processing: Wavelets, Curvelets, Morphological Diversity*. Cambridge Univ. Press, Cambridge
Stewart I. M., Fenech D. M., Muxlow T. W. B., 2011, *A&A*, 535, A81
Tang M., Repetti A., 2023, preprint (arXiv:2304.11200)
Terris M., Dabbech A., Tang C., Wiaux Y., 2022, *MNRAS*, 518, 604
The CASA Team, 2022, *PASP*, 134, 114501
Thompson A. R., Moran J. M., Swenson George W. J., 2017, *Interferometry and Synthesis in Radio Astronomy*, 3rd edn. Cambridge Univ. Press, Cambridge
Thouvenin P.-A., Abdulaziz A., Dabbech A., Repetti A., Wiaux Y., 2022a, *MNRAS*, 521, 1
Thouvenin P.-A., Dabbech A., Jiang M., Abdulaziz A., Thiran J.-P., Jackson A., Wiaux Y., 2022b, *MNRAS*, 521, 20
Tiede P., 2022, *J. Open Source Softw.*, 7, 4457

Tse D., Viswanath P., 2005, *Fundamentals of Wireless Communication*. Cambridge Univ. Press, Cambridge
van der Walt S. et al., 2014, *PeerJ*, 2, e453
Venkatakrisnan S. V., Bouman C. A., Wohlberg B., 2013, in *IEEE Global Conference on Signal and Information Processing*, p. 945
Vidal A. F., De Bortoli V., Pereyra M., Durmus A., 2020, *SIAM J. Imaging Sci.*, 13, 1945
Wang R., Chen Z., Luo Q., Wang F., 2023, preprint (arXiv:2305.09121)
Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vanderghyest P., 2009, *MNRAS*, 395, 1733
Wilber A. G., Dabbech A., Terris M., Jackson A., Wiaux Y., 2023, *MNRAS*, 522, 5576
Xu K., Ge H., Tebbutt W., Tarek M., Trapp M., Ghahramani Z., 2020, in Zhang C., Ruiz F., Bui T., Dieng A. B., Liang D., eds, *Proc. Machine Learning Research Vol. 118, Proceedings of the 2nd Symposium on Advances in Approximate Bayesian Inference*. PMLR, p. 1. Available at: <https://proceedings.mlr.press/v118/xu20a.html>
Yu J., Fan Y., Yang J., Xu N., Wang Z., Wang X., Huang T., 2018, preprint (arXiv:1808.08718)

APPENDIX A: MAP CALCULATION FOR THE WAVELET-BASED PRIOR MODEL

The MAP estimation for the wavelet-based model can be recasted as the following optimization problem

$$\hat{\mathbf{x}}_{\text{MAP}} = \underset{\mathbf{x} \in \mathbb{R}^N}{\text{argmin}} \frac{1}{2\sigma^2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 + \lambda_{\text{wav}} \|\Psi^\dagger \mathbf{x}\|_1 + t_{\mathbb{R}^N}(\mathbf{x}), \quad (\text{A1})$$

where λ_{wav} corresponds to the regularization strength of the wavelet-based prior. The FISTA algorithm to estimate the MAP is presented in Algorithm 2 where we use the soft thresholding operator, $\text{soft}(\cdot)$, defined in equation (49). The Lipschitz constant used to define the step size can be set as $L_{\text{wav}} = \|\Phi^\dagger \Phi\|/\sigma^2$.

Algorithm 2: FISTA (Beck & Teboulle) tackling (25)

- 1 **Input:** $\Psi, \Phi, \sigma, \lambda_{\text{wav}}, \xi, a_{(1)} = 1, \mathbf{z}_{(1)} = \mathbf{x}_{(0)} = \text{Re}(\Phi^\dagger \mathbf{y}),$
 $\tau_{\text{wav}} = 0.98/L_{\text{wav}}.$
 - 2 **Output:** $\hat{\mathbf{x}}_{\text{MAP}}$
 - 3 **for** $n = 1, \dots, N_{\text{max}}$ **do**
 - 4 $\tilde{\mathbf{z}}_{(n)} = \mathbf{z}_{(n)} - \frac{\tau_{\text{wav}}}{\sigma^2} \text{Re}(\Phi^\dagger(\Phi\mathbf{z}_{(n)} - \mathbf{y}))$
 - 5 $\mathbf{x}_{(n)} = \tilde{\mathbf{z}}_{(n)} + \Psi \left(\text{soft}_{\lambda_{\text{wav}}\tau}(\Psi^\dagger \tilde{\mathbf{z}}_{(n)}) - \Psi^\dagger \tilde{\mathbf{z}}_{(n)} \right)$
 - 6 $a_{(n+1)} = \frac{1}{2} \left(1 + \sqrt{4a_{(n)}^2 + 1} \right)$
 - 7 $\mathbf{z}_{(n+1)} = \mathbf{x}_{(n)} + \frac{a_{(n)} - 1}{a_{(n+1)}} (\mathbf{x}_{(n)} - \mathbf{x}_{(n-1)})$
 - 8 **if** $\frac{\|\mathbf{x}_{(n)} - \mathbf{x}_{(n-1)}\|}{\|\mathbf{x}_{(n-1)}\|} < \xi$ **then**
 - 9 | **break**
 - 10 **end**
 - 11 **end**
 - 12 **set** $\hat{\mathbf{x}}_{\text{MAP}} = \mathbf{x}_{(n)}$
-

APPENDIX B: WAVELET-BASED BAYESIAN UNCERTAINTY QUANTIFICATION

B1 Hypothesis testing of image structure

Fig. B1 and Table B1 present the results of the hypothesis testing of structure for the model with the sparsity-promoting prior.

Table B1. Hypothesis test results for the inpainted surrogates from Fig. B1 generated with the wavelet-based model. All values are in units 10^4 . The description of Table 2 applies in this table.

Images	Test area	Ground truth	Method	Initial $(f + g)(\hat{x}^*)$	Surrogate $(f + g)(\hat{x}^{*,sgt})$	Isocontour $\hat{\gamma}_{0.01}$	Hypothesis test
M31	1	✓	SK-ROCK	0.448	1.396	1.105	✓
			MAP	0.359	1.335	1.039	✓
Cygnus A	1	✓	SK-ROCK	0.480	0.533	1.639	✗
			MAP	0.444	0.514	1.789	✗
W28	1	✓	SK-ROCK	0.353	5.190	0.879	✓
			MAP	0.284	5.204	0.964	✓
3C288	1	✓	SK-ROCK	0.729	2.487	1.398	✓
			MAP	0.654	2.409	1.333	✓
	2	✗	SK-ROCK	0.729	0.729	1.398	✗
			MAP	0.654	0.654	1.333	✗

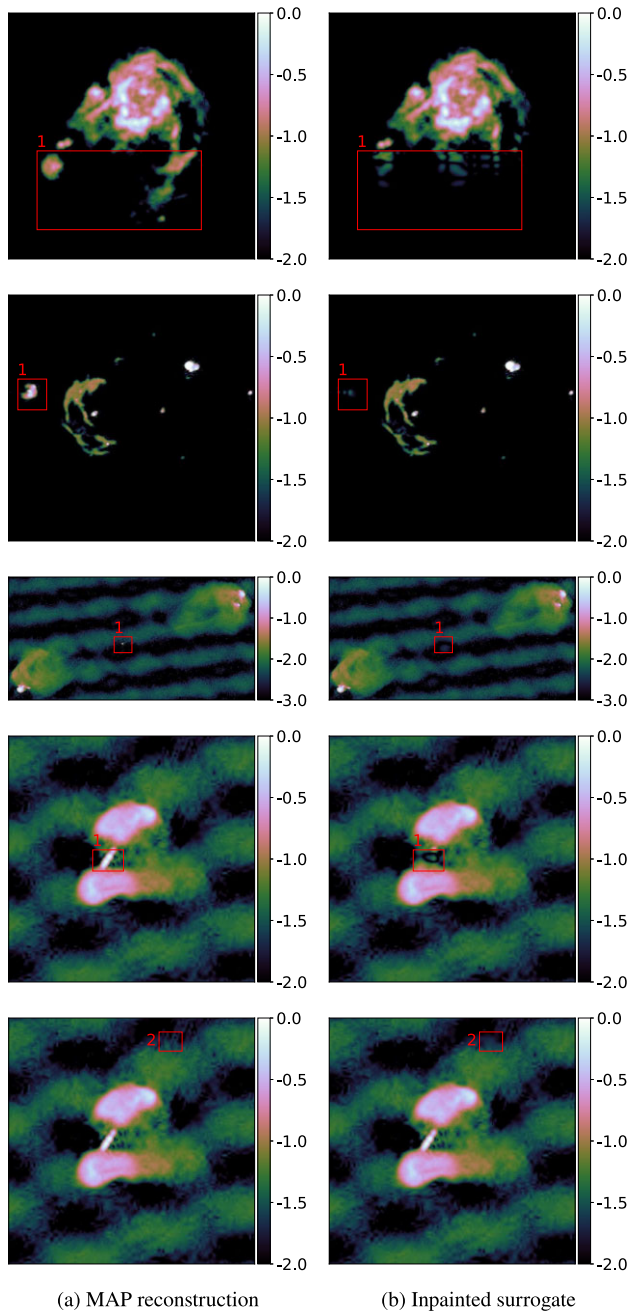


Figure B1. Hypothesis test of different regions of the four images, M31, W28, Cygnus A, and 3C288. All the images are shown in \log_{10} scale. The figure is similar to Fig. 5, but the wavelet-based model has been used to generate the MAP. The wavelet prior was used to inpaint the surrogate image.

APPENDIX C: MORE REALISTIC EXPERIMENT RESULTS

Figs C1, C2, and C3 present the results using the realistic MeerKAT ungridded visibility pattern of the images W28, Cygnus A, and 3C288, correspondingly. Tables C1, C2, and C3 show the quantitative results of the experiment.

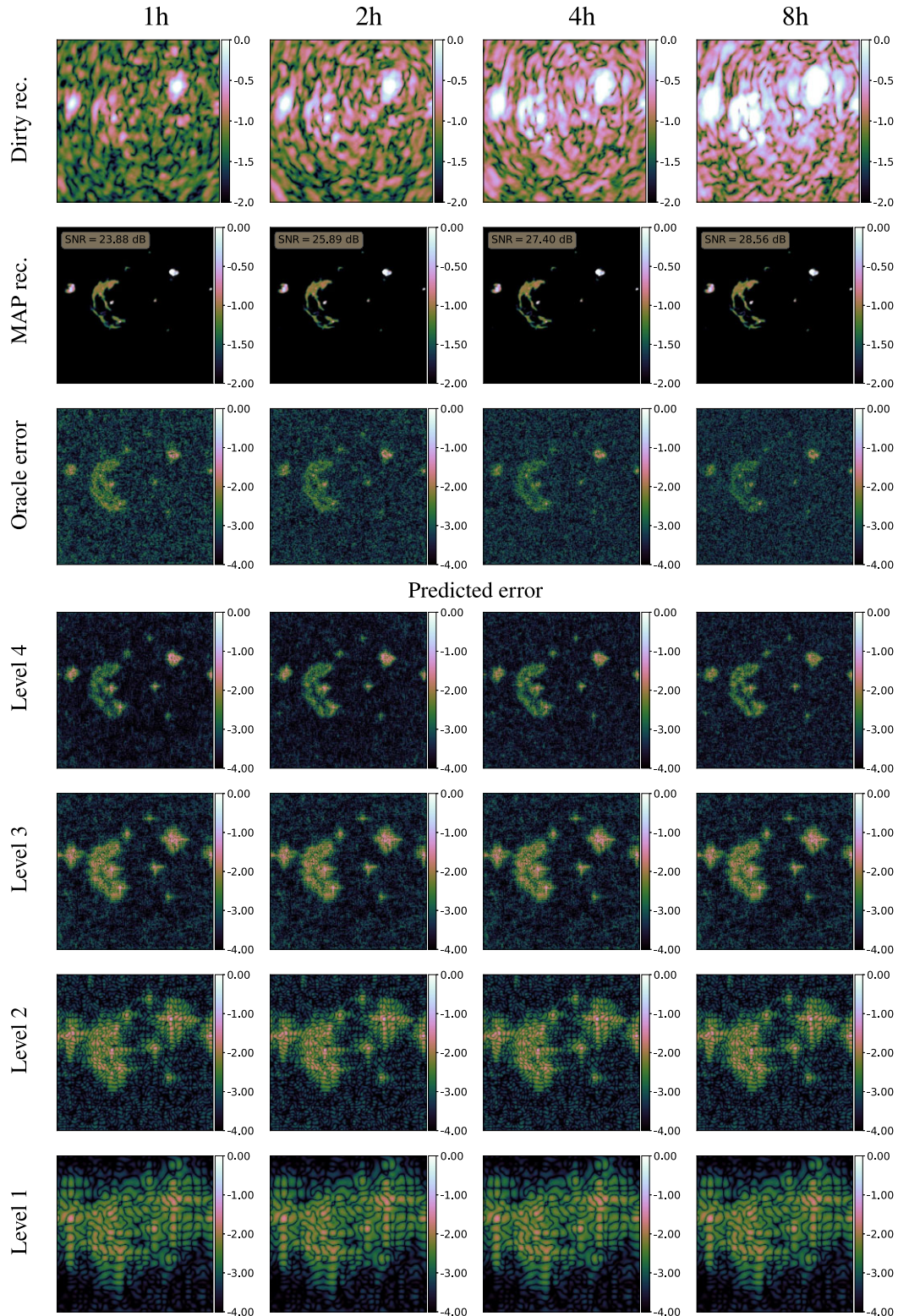


Figure C1. Reconstructions and fast pixel uncertainty quantification (UQ) with the QUANTIFAI model for the W28 image with the four sets of simulated MeerKAT ungridded visibilities with a field of view of approximately 1 deg^2 . Each column corresponds to the four data sets with synthesis times of 1, 2, 4, and 8 h. The first row represents the dirty reconstruction. The MAP reconstruction is presented in the second row, while the oracle error, which we do not have access to with real data, is shown in the third row. The different decomposition levels of pixel UQ are shown in the last four rows.

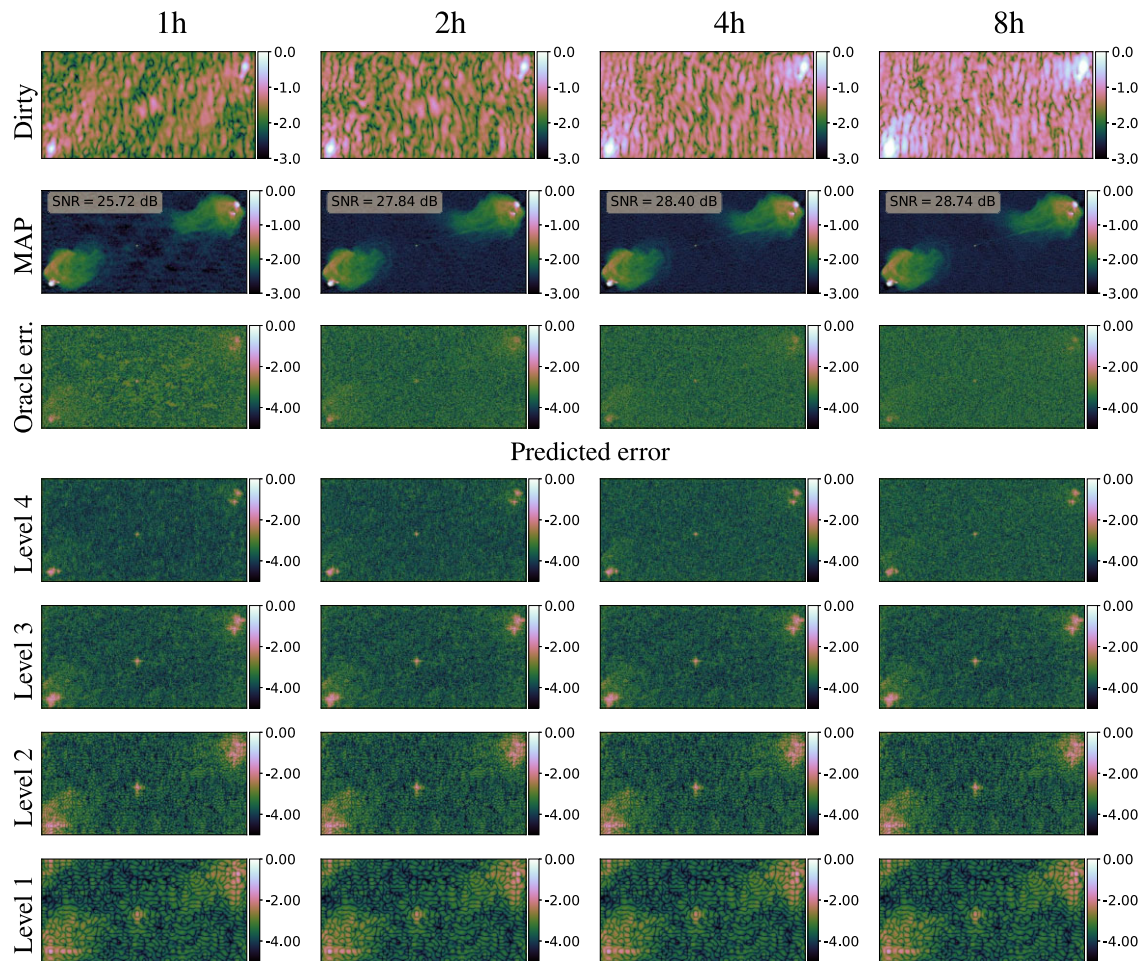


Figure C2. Reconstructions and fast pixel uncertainty quantification (UQ) with the QUANTIFAI model for the Cygnus A image with the four sets of simulated MeerKAT ungridded visibilities with a field of view of approximately 1 deg^2 . Each column corresponds to the four data sets with synthesis times of 1, 2, 4, and 8 h. The first row represents the dirty reconstruction. The MAP reconstruction is presented in the second row, while the oracle error, which we do not have access to with real data, is shown in the third row. The different decomposition levels of pixel UQ are shown in the last four rows.

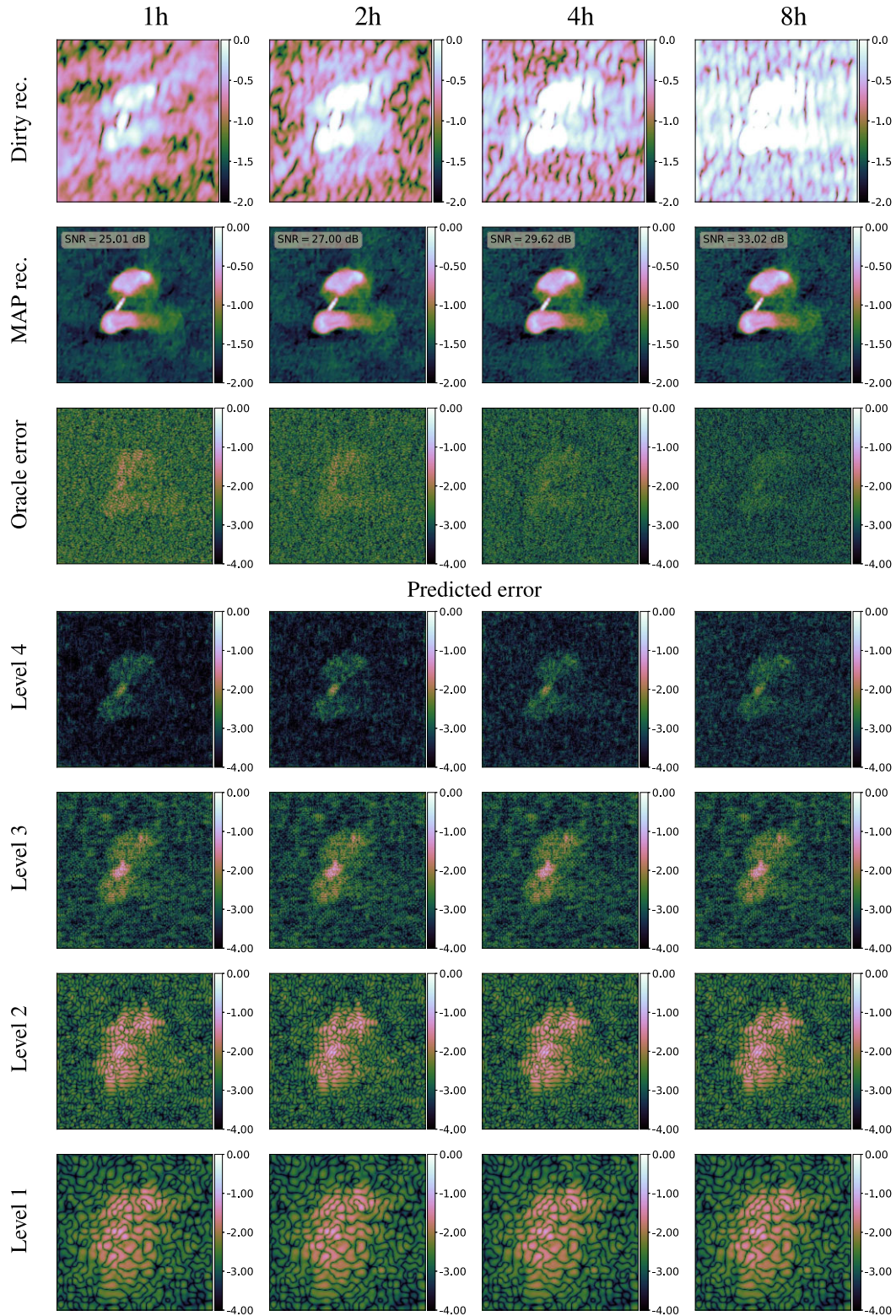


Figure C3. Reconstructions and fast pixel uncertainty quantification (UQ) with the QUANTIFAI model for the 3C288 image with the four sets of simulated MeerKAT ungridded visibilities with a field of view of approximately 1 deg^2 . Each column corresponds to the four data sets with synthesis times of 1, 2, 4, and 8 h. The first row represents the dirty reconstruction. The MAP reconstruction is presented in the second row, while the oracle error, which we do not have access to with real data, is shown in the third row. The different decomposition levels of pixel UQ are shown in the last four rows.

Table C1. Main results of QUANTIFAI for the W28 image with the realistic ungridded MeerKAT visibility patterns with differing synthesis times. As the number of visibilities grows with the synthesis timer, so does the reconstruction SNR. The number of visibilities increases proportionally to the synthesis times.

Metrics		Data sets			
		1 h	2 h	4 h	8 h
Reconstruction	Number of visibilities	3×10^4	6×10^4	1.2×10^5	2.4×10^5
	MAP reconstruction SNR (dB)	23.88	25.89	27.40	28.56
	Measurement op. evaluations	6976	6124	5270	4062
	Wall-clock time (s)	34.83	59.25	93.61	137.2
UQ	Measurement op. evaluations	30	30	32	32
	Wall-clock time (s)	0.31	0.47	0.78	1.35

Table C2. Main results of QUANTIFAI for the Cygnus A image with the realistic ungridded MeerKAT visibility patterns with differing synthesis times. As the number of visibilities grows with the synthesis timer, so does the reconstruction SNR. The number of visibilities increases proportionally to the synthesis times.

Metrics		Data sets			
		1 h	2 h	4 h	8 h
Reconstruction	Number of visibilities	3×10^4	6×10^4	1.2×10^5	2.4×10^5
	MAP reconstruction SNR (dB)	25.72	27.84	28.40	28.74
	Measurement op. evaluations	6482	5172	3686	2692
	Wall-clock time (s)	36.17	51.71	66.79	91.79
UQ	Measurement op. evaluations	30	30	30	32
	Wall-clock time (s)	0.35	0.50	0.77	1.36

Table C3. Main results of QUANTIFAI for the 3C288 image with the realistic ungridded MeerKAT visibility patterns with differing synthesis times. As the number of visibilities grows with the synthesis timer, so does the reconstruction SNR. The number of visibilities increases proportionally to the synthesis times.

Metrics		Data sets			
		1 h	2 h	4 h	8 h
Reconstruction	Number of visibilities	3×10^4	6×10^4	1.2×10^5	2.4×10^5
	MAP reconstruction SNR (dB)	25.01	27.00	29.62	33.02
	Measurement op. evaluations	2730	2198	1976	1856
	Wall-clock time (s)	13.93	21.22	34.99	62.68
UQ	Measurement op. evaluations	26	28	32	32
	Wall-clock time (s)	0.28	0.44	0.78	1.33

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.