Uncertainty quantification for radio interferometric imaging: II. MAP estimation

Xiaohao Cai[1★], Marcelo Pereyra[2★] and Jason D. McEwen[1★]
[1]*Mullard Space Science Laboratory, University College London (UCL), Surrey RH5 6NT, United Kingdom*
[2]*Maxwell Institute for Mathematical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, United Kingdom*

Accepted —. Received —; in original form —

ABSTRACT
Uncertainty quantification is a critical missing component in radio interferometric imaging that will only become increasingly important as the big-data era of radio interferometry emerges. Statistical sampling approaches to perform Bayesian inference, like Markov Chain Monte Carlo (MCMC) sampling, can in principle recover the full posterior distribution of the image, from which uncertainties can then be quantified. However, for massive data sizes, like those anticipated from the Square Kilometre Array (SKA), it will be difficult if not impossible to apply any MCMC technique due to its inherent computational cost. We formulate Bayesian inference problems with sparsity-promoting priors (motivated by compressive sensing), for which we recover *maximum a posteriori* (MAP) point estimators of radio interferometric images by convex optimisation. Exploiting recent developments in the theory of probability concentration, we quantify uncertainties by post-processing the recovered MAP estimate. Three strategies to quantify uncertainties are developed: (i) highest posterior density credible regions; (ii) local credible intervals (*cf.* error bars) for individual pixels and superpixels; and (iii) hypothesis testing of image structure. These forms of uncertainty quantification provide rich information for analysing radio interferometric observations in a statistically robust manner. Our MAP-based methods are approximately 10^5 times faster computationally than state-of-the-art MCMC methods and, in addition, support highly distributed and parallelised algorithmic structures. For the first time, our MAP-based techniques provide a means of quantifying uncertainties for radio interferometric imaging for realistic data volumes and practical use, and scale to the emerging big-data era of radio astronomy.

Key words: techniques: image processing – techniques: interferometric – methods: data analysis – methods: numerical – methods: statistical.

1 INTRODUCTION

Radio interferometric (RI) telescopes provide observations of the radio emission of the sky with high angular resolution and sensitivity, and provide a wealth of valuable information for astrophysics and cosmology (Ryle & Vonberg 1946; Ryle & Hewish 1960; Thompson et al. 2017). Radio interferometers essentially acquire Fourier measurements of the sky image of interest. Imaging observations made by radio interferometers thus requires solving an ill-posed linear inverse problem (Thompson et al. 2017), which is an important first step in many subsequent scientific analyses. Since the inverse problem is ill-posed (sometimes seriously), uncertainty information (*e.g.* error estimates) regarding reconstructed images is critical. Nevertheless, uncertainty information is currently lacking in all RI imaging techniques used in practice. In Cai et al. (2017a), the first of these companion articles, we propose uncertainty quantification strategies for RI imaging based on state-of-the-art Markov chain Monte Carlo (MCMC) methods that sample the full posterior distribution of the image, with the sparsity-promoting priors that have been shown in practice to be highly effective (*e.g.* Pratley et al. 2018). Excellent results were achieved and a variety of different uncertainty quantification strategies were presented. However, it is difficult to scale these strategies to big-data due to their high computational overhead. We address this issue in the current article.

Over the coming decades radio astronomy will transition into the so-called big-data era. Generally speaking, the new generation of radio telescopes, such as the LOw Frequency ARray (LOFAR[1]), the Extended Very Large Array (EVLA[2]), the Australian

★ E-mail: x.cai@ucl.ac.uk (XC); m.pereyra@hw.ac.uk (MP); jason.mcewen@ucl.ac.uk (JDM)

[1] http://www.lofar.org
[2] http://www.aoc.nrao.edu/evla

Square Kilometre Array Pathfinder (ASKAP[3]), and the Murchison Widefield Array (MWA[4]), will achieve much higher dynamic range and angular resolution than previous instruments and will acquire very large volumes of data. The Square Kilometer Array (SKA[5]) will provide a considerable step again in dynamic range (six or seven orders of magnitude beyond prior telescopes) and angular resolution, and will acquire massive volumes of data, ushering in the big-data era of radio astronomy. This emerging era of big-data, inevitably, will bring further challenges and so uncertainty quantification will be increasingly important. As discussed in Cai et al. (2017a), existing image reconstruction techniques, such as CLEAN-based methods (Högbom 1974; Bhatnagar & Corwnell 2004; Cornwell 2008; Stewart et al. 2011), the maximum entropy method (MEM) (Ables 1974; Gull & Daniell 1978; Cornwell & Evans 1985), and compressed sensing (CS) methods (Wiaux et al. 2009a,b; Suksmono 2009; Wenger et al. 2010; McEwen & Wiaux 2011; Li et al. 2011a,b; Carrillo et al. 2012, 2014; Wolz et al. 2013; Dabbech et al. 2015; Dabbech et al. 2017; Garsden et al. 2015; Onose et al. 2016, 2017; Pratley et al. 2018; Kartik et al. 2017), do not provide uncertainty information regarding their reconstructed images. The approaches that do provide some form of uncertainty quantification (Sutter et al. 2014; Junklewitz et al. 2016; Greiner et al. 2017) cannot scale to big-data due to their high computational cost, are typically restricted to Gaussian or log-normal priors, and are not currently used in practice. Please see our first article in this companion series (Cai et al. 2017a) for a more thorough review of RI imaging techniques and their properties.

The current state of the field thus triggers an urgent need to develop efficient uncertainty quantification methods for RI imaging that scale to big-data. Furthermore, we seek to support the sparsity-promoting priors that have been demonstrated in practice to be highly effective for RI imaging (*e.g.* Pratley et al. 2018). In Cai et al. (2017a) (the first part of this companion series), we proposed uncertainty quantification methods to address the RI imaging problem with sparse priors. In the current article (the second part of this companion series), we present fast uncertainty quantification methods that not only support sparse priors but also scale to big-data. The techniques presented in this article are very different to those presented in Cai et al. (2017a) but support the same forms of uncertainty quantification.

The uncertainty quantification methods proposed in Cai et al. (2017a) are based on two proximal MCMC sampling methods, *i.e.* the Moreau-Yoshida unadjusted Langevin algorithm (MYULA) (Durmus et al. 2016) and the proximal Metropolis-adjusted Langevin algorithm (Px-MALA) (Pereyra 2016b). The main steps of the uncertainty quantification strategies presented in Cai et al. (2017a) can be briefly summarised as follows: firstly, the posterior distribution of the image is MCMC sampled; then, uncertainty quantification is performed by using the generated samples to compute local (pixel-wise) credible intervals, highest posterior density (HPD) credible regions, and to perform hypothesis testing of image structure. Two frameworks – analysis and synthesis models – are considered. While excellent results were achieved in Cai et al. (2017a), when it comes to big-data, the proposed approach would suffer due to the long computation time required to sample the posterior distribution (as would be the case for any MCMC sampling approach).

[3] http://www.atnf.csiro.au/projects/askap
[4] http://www.mwatelescope.org/telescope
[5] http://www.skatelescope.org/

In this article we exploit an analytic method to approximate HPD credible regions from *maximum a posteriori* (MAP) estimators, as derived in Pereyra (2017), in order to develop very fast methods to perform uncertainty quantification for RI imaging. Our approach supports sparse priors and scales to massive data sizes, *i.e.* to big-data. We begin by formulating Bayesian MAP estimation for RI imaging as unconstrained convex optimisation problems, for analysis and synthesis forms. These are subsequently solved efficiently by using convex minimisation algorithms (*e.g.* Combettes & Pesquet 2010). Recent advances in convex optimisation have resulted in techniques that achieve excellent reconstruction fidelity (with convergence guarantees), are flexible, and exhibit relatively low computational costs. They also afford algorithmic structures that can be highly distributed and parallelised (*e.g.* Carrillo et al. 2014; Onose et al. 2016) and computed in an online manner (Cai et al. 2017b). Note, specifically, that only one point estimator is computed here for the analysis or synthesis form, in contrast to sampling approaches that seek to explore the full posterior distribution as in Cai et al. (2017a), which is very time consuming. MAP estimation is then followed by various strategies to quantify uncertainties. Precisely, first the method of Pereyra (2017) is used to obtain approximate HPD credible regions for the recovered image. These HPD regions are then used, for the first time, to compute local credible intervals (*cf.* error bars) that analyse uncertainty spatially and at different scales (pixles or superpixels). Finally, we also use the HPD credible regions to perform hypothesis tests of image structure. We test our proposed approaches on simulated RI observations to demonstrate their effectiveness and compare with the MCMC methods presented in Cai et al. (2017a).

The remainder of this article is organised as follows. In Section 2 we review the RI imaging inverse problem. In Section 3 we apply convex optimisation algorithms to solve the MAP estimation problem for RI imaging in the context of sparse priors. Note that Sections 2 and 3 review background material for our specific problem to provide clarity and completeness (*i.e.* so that all derivations are explicit and thus one could implement our methods if one wanted). Uncertainty quantification techniques for RI imaging based on MAP estimation are formulated in Section 4. The performance of the proposed methods is then evaluated numerically in Section 5, where we compare uncertainties quantified by proximal MCMC methods and by MAP estimation. Finally, we conclude in Section 6 with a summary of our main contributions and a discussion of planned extensions.

2 RADIO INTERFEROMETRIC IMAGING

In this section the inverse problem related to RI image reconstruction is introduced. We briefly recall the use of proximal MCMC methods to solve this problem (Cai et al. 2017a), which we use as a benchmark in the experiments that follow. Finally, an introduction to Bayesian MAP estimation approaches for RI imaging is presented, which may be solved by efficient convex optimisation strategies.

2.1 Radio interferometry

Here, we concisely recall the inverse problem of RI imaging (for further details see Cai et al. 2017a and references therein).

In the discretised setting, let $\boldsymbol{x} \in \mathbb{R}^N$ represent the sampled intensity signal (the sky brightness distribution). In particular, $\boldsymbol{x}$

can be represented by

$$\boldsymbol{x} = \boldsymbol{\Psi}\boldsymbol{a} = \sum_i \boldsymbol{\Psi}_i a_i, \tag{1}$$

where $\boldsymbol{\Psi} \in \mathbb{C}^{N\times L}$ is a basis or dictionary (*e.g.*, a wavelet basis or an over-complete frame) and vector $\boldsymbol{a} = (a_1, \cdots, a_L)^\top$ represents the synthesis coefficients of $\boldsymbol{x}$ under $\boldsymbol{\Psi}$. In particular, $\boldsymbol{x}$ is said to be sparse if $\boldsymbol{a}$ contains only K non-zero coefficients, $K \ll N$, or compressible if many coefficients of $\boldsymbol{a}$ are nearly zero. In practice, it is ubiquitous that natural images are sparse or compressible for approriate choices of $\boldsymbol{\Psi}$. Refer to Cai et al. (2017a) for more details about sparse representation.

Let $\boldsymbol{y} \in \mathbb{C}^M$ be the M visibilities acquired by a radio interferometric telescope observed under a linear measurement operator $\boldsymbol{\Phi} \in \mathbb{C}^{M\times N}$ modelling the acquisition of the sky brightness distribution. Then, we have

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{n}, \tag{2}$$

where $\boldsymbol{n} \in \mathbb{C}^M$ is the instrumental noise. Without loss of generality, we subsequently consider independent and identically distributed (i.i.d.) Gaussian noise. In practice, $\boldsymbol{y}$ is only observed partially or with limited resolution. Recovering the sky intensity signal $\boldsymbol{x}$ from the measured visibilities $\boldsymbol{y}$ acquired according to equation (2) then amounts to solving a linear inverse problem (Rau et al. 2009).

2.2 Bayesian inference

The RI inverse problem (2) can be solved elegantly in the Bayesian statistical framework, which provides tools to estimate $\boldsymbol{x}$ (or $\boldsymbol{a}$) as well as to quantify the uncertainty in the estimated solutions. After combining the observed and prior information, the posterior distribution $p(\boldsymbol{x}|\boldsymbol{y})$ (or $p(\boldsymbol{a}|\boldsymbol{y})$) can be obtained by using Bayes' theorem. Refer to Cai et al. (2017a) for more detailed discussion about Bayesian inference in the context of RI imaging.

2.3 Proximal MCMC methods

To solve the ill-posed inverse problem in (2) with sparsity-promoting priors, which have been shown in practice to be highly effective (Pratley et al. 2018), while also performing uncertainty quantification, two proximal MCMC methods to perform Bayesian inference for RI imaging were developed in the companion article (Cai et al. 2017a). These proximal MCMC methods seek to sample the full posterior density $p(\boldsymbol{x}|\boldsymbol{y})$ that models our understanding of the image $\boldsymbol{x}$ given data $\boldsymbol{y}$, in the context of prior information. From the full posterior, summary estimators of $\boldsymbol{x}$ and other quantities of interest can be computed. In particular, in Cai et al. (2017a) these methods are used to perform a range of uncertainty quantification analysis for RI images.

One of the proximal MCMC methods presented in Cai et al. (2017a), MYULA, scales efficiently to high dimensions but suffers from some estimation bias (Durmus et al. 2016). The other, Px-MALA, corrects this bias by using a Metropolis-Hastings correction step, at the expense of a higher computational cost and slower convergence (Pereyra 2016b). Since Px-MALA can provide results with corrected bias and thus is more accurate, we use it as a benchmark in the subsequent numerical tests presented in this work. Nevertheless, the MCMC methods discussed in Cai et al. (2017a) will suffer when scaling to big-data (as will any MCMC method), which motives us to explore alternative faster methods that can scale to big-data.

In this article we develop methods for uncertainty quantification based on MAP estimation. We emphasise that while MCMC methods such as Px-MALA are not as efficient as MAP estimation (the main focus in this article), and do not scale to large RI datasets, they are useful for smaller datasets and as a benchmark for the efficient alternative methods that we propose in Section 4.

2.4 Maximum a posteriori (MAP) estimation

As discussed in the previous sections, sampling the full posterior $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$ by MCMC methods is difficult because of the high dimensionality involved. Instead, Bayesian estimators that summarise $p(\boldsymbol{x}|\boldsymbol{y})$ or $p(\boldsymbol{a}|\boldsymbol{y})$ are often computed. In particular, one common approach is to compute MAP (maximum-a-posteriori) estimators given by

$$\boldsymbol{x}_{\rm map} = \underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ \mu \|\boldsymbol{\Psi}^\dagger \boldsymbol{x}\|_1 + \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2 / 2\sigma^2 \right\}, \tag{3}$$

for the analysis model, and for the synthesis model by

$$\boldsymbol{x}_{\rm map} = \boldsymbol{\Psi} \times \underset{\boldsymbol{a}}{\operatorname{argmin}} \left\{ \mu \|\boldsymbol{a}\|_1 + \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2 / 2\sigma^2 \right\}, \tag{4}$$

where the first term is a prior distribution to regularise the problem, reduce uncertainty, and improve estimation results, and the second term is associated with the likelihood function of the model associated with (2).

As we discuss below, a main computational advantage of the MAP estimators (3) and (4) is that they can be computed very efficiently, even in high dimensions, by using convex optimisation algorithms (*e.g.* Combettes & Pesquet 2010; Green et al. 2015). There is also abundant empirical evidence suggesting that these estimators deliver accurate reconstruction results (see Pereyra 2016a also for a theoretical analysis of MAP estimation). However, since MAP estimation results in a single point estimator, we typical lose uncertainty information that MCMC methods can provide (Cai et al. 2017a). On the contrary, however, as we show in this article it is possible to approximately quantify the uncertainties associated with MAP estimators by leveraging recent results in the theory of probability concentration (Pereyra 2017). Consequently, using the techniques presented later in this article MAP estimation can provide fast methods that scale to big-data and that quantify uncertainties.

2.5 Convex optimisation methods for MAP estimation

There are several convex optimisation methods that can be used to solve the MAP estimation problems (3) and (4) efficiently, such as forward-backward splitting, Douglas-Rachford splitting, or alternating direction method of multipliers (ADMM) (see Combettes & Pesquet 2010). In our experiments (3) and (4) are solved by adopting the simple forward-backward algorithm, which we detail in Appendix A.

3 SPARSE MAP ESTIMATION FOR RI IMAGING

In this section we present the algorithmic details of implementing the forward-backward splitting algorithm to solve the sparse MAP estimation problems for both the analysis setting (3) and synthesis setting (4). For the sake of brevity, henceforth the labels $^-$ and $\hat{}$ denote symbols related to the analysis and synthesis models, respectively.

3.1 Analysis

For the analysis setting (3), set $\bar{f}(\boldsymbol{x}) = \mu\|\boldsymbol{\Psi}^\dagger\boldsymbol{x}\|_1$ and $\bar{g}(\boldsymbol{x}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{x}\|_2^2/2\sigma^2$. Then

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ \bar{f}(\boldsymbol{x}) + \bar{g}(\boldsymbol{x}) \right\} \tag{5}$$

can be solved using the forward-backward iteration formula (shown in Appendix A), leading to the iterations

$$\boldsymbol{x}^{(i+1)} = \operatorname{prox}_{\lambda^{(i)}\bar{f}}(\boldsymbol{x}^{(i)} - \lambda^{(i)}\nabla\bar{g}(\boldsymbol{x}^{(i)})). \tag{6}$$

Assume for now $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} = \mathsf{I}$, where I is identity matrix (although this assumption is not essential and relaxed later). We have, $\forall \bar{\boldsymbol{z}} \in \mathbb{R}^N$,

$$\operatorname{prox}_{\lambda\bar{f}}(\bar{\boldsymbol{z}}) = \bar{\boldsymbol{z}} + \boldsymbol{\Psi}\left(\operatorname{soft}_{\lambda\mu}(\boldsymbol{\Psi}^\dagger\bar{\boldsymbol{z}}) - \boldsymbol{\Psi}^\dagger\bar{\boldsymbol{z}}\right), \tag{7}$$

and

$$\nabla\bar{g}(\boldsymbol{x}) = \boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x} - \boldsymbol{y})/\sigma^2, \tag{8}$$

where $\operatorname{soft}_{\lambda\mu}(\boldsymbol{z})$ is the pointwise soft-thresholding operator of vector $\boldsymbol{z}$ defined in (A5). See Remark 4.1 in Cai et al. (2017a) when $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} \neq \mathsf{I}$ for computing $\operatorname{prox}_{\lambda\bar{f}}(\bar{\boldsymbol{z}})$. Substituting (7) and (8) into (6), the analysis problem (3) can be solved iteratively by

$$\boldsymbol{v}^{(i+1)} = \boldsymbol{x}^{(i)} - \lambda^{(i)}\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x}^{(i)} - \boldsymbol{y})/\sigma^2, \tag{9}$$

$$\boldsymbol{x}^{(i+1)} = \boldsymbol{v}^{(i+1)} + \boldsymbol{\Psi}\left(\operatorname{soft}_{\lambda^{(i)}\mu}(\boldsymbol{\Psi}^\dagger\boldsymbol{v}^{(i+1)}) - \boldsymbol{\Psi}^\dagger\boldsymbol{v}^{(i+1)}\right). \tag{10}$$

As initialisation use, *e.g.*, $\boldsymbol{x}^{(0)} = \boldsymbol{\Phi}^\dagger\boldsymbol{y}$, *i.e.* the dirty image.

3.2 Synthesis

For the synthesis setting (4), set $\hat{f}(\boldsymbol{a}) = \mu\|\boldsymbol{a}\|_1$ and $\hat{g}(\boldsymbol{a}) = \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}\|_2^2/2\sigma^2$. Then

$$\underset{\boldsymbol{x}}{\operatorname{argmin}} \left\{ \hat{f}(\boldsymbol{a}) + \hat{g}(\boldsymbol{a}) \right\} \tag{11}$$

can be solved using the forward-backward iteration formula (shown in Appendix A), leading to the iterations

$$\boldsymbol{a}^{(i+1)} = \operatorname{prox}_{\lambda^{(i)}\hat{f}}(\boldsymbol{a}^{(i)} - \lambda^{(i)}\nabla\hat{g}(\boldsymbol{a}^{(i)})). \tag{12}$$

We have, $\forall \hat{\boldsymbol{z}} = (\hat{z}_1, \cdots, \hat{z}_L) \in \mathbb{R}^L$,

$$\begin{aligned} \operatorname{prox}_{\lambda\hat{f}}(\hat{\boldsymbol{z}}) &= \underset{\boldsymbol{u}\in\mathbb{R}^L}{\operatorname{argmin}}\, \lambda\mu\|\boldsymbol{u}\|_1 + \|\boldsymbol{u} - \hat{\boldsymbol{z}}\|^2/2 \\ &= \operatorname{soft}_{\lambda\mu}(\hat{\boldsymbol{z}}) \end{aligned} \tag{13}$$

and

$$\nabla\hat{g}(\boldsymbol{a}) = \boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a} - \boldsymbol{y})/\sigma^2. \tag{14}$$

Finally, substituting (13) and (14) into (12), the synthesis problem (4) can be solved iteratively by

$$\boldsymbol{a}^{(i+1)} = \operatorname{soft}_{\lambda^{(i)}\mu}\left(\boldsymbol{a}^{(i)} - \lambda^{(i)}\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}^{(i)} - \boldsymbol{y})/\sigma^2\right). \tag{15}$$

Remark 3.1. Note that in both the analysis and synthesis settings various terms can be precomputed. For example, in (9) and (14) the operators $\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger\boldsymbol{\Phi}\boldsymbol{\Psi}$ can be precomputed offline. Similarly, the terms of $\boldsymbol{\Phi}^\dagger\boldsymbol{y}$ (the so-called dirty map) and $\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger\boldsymbol{y}$ respectively in (9) and (14) can also be precomputed to improve computation efficiency.

We summarise the forward-backward splitting algorithms for the analysis and synthesis reconstruction forms in Algorithms 1 and 2. We consider stopping criteria based on a maximum iteration number and when the relative difference between solutions at two consecutive iterations is within some tolerance, *i.e.*, $\|\boldsymbol{x}^{(i+1)} - \boldsymbol{x}^{(i)}\|_2/\|\boldsymbol{x}^{(i)}\|_2$ (for Algorithm 1) and $\|\boldsymbol{\Psi}\boldsymbol{a}^{(i+1)} - \boldsymbol{\Psi}\boldsymbol{a}^{(i)}\|_2/\|\boldsymbol{\Psi}\boldsymbol{a}^{(i)}\|_2$ (for Algorithm 2). The iteration is terminated when either of the stopping criteria are reached. The complexity of the algorithms is simply given by the complexity of application of the measurement operator $\boldsymbol{\Phi}$. However, the measurement operator (and its adjoint) needs to be applied multiple times, hence the pre-factor associated with the complexity is significant. In general fast, optimised algorithms are applied for realistic measurement operators (essentially based on non-uniform fast Fourier transforms), resulting in a complexity of $\mathcal{O}(MJ + N\log N)$, where J denotes the support of the kernel used to perform convolutional degridding (see, *e.g.*, Pratley et al. 2018 for further details).

Algorithm 1: Forward-backward algorithm for analysis

Input: $\boldsymbol{y} \in \mathbb{R}^M$, $\boldsymbol{x}^{(0)} \in \mathbb{R}^N$, σ and $\lambda^{(i)} \in (0, \infty)$
Output: $\boldsymbol{x}'$
do
  update $\boldsymbol{v}^{(i+1)} = \boldsymbol{x}^{(i)} - \lambda^{(i)}\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{x}^{(i)} - \boldsymbol{y})/\sigma^2$
  compute $\boldsymbol{u} = \boldsymbol{\Psi}^\dagger\boldsymbol{v}^{(i+1)}$
  update $\boldsymbol{x}^{(i+1)} = \boldsymbol{v}^{(i+1)} + \boldsymbol{\Psi}\left(\operatorname{soft}_{\lambda^{(i)}\mu}(\boldsymbol{u}) - \boldsymbol{u}\right)$
  $i = i + 1$
while *Stopping criterion is not reached*;
set $\boldsymbol{x}' = \boldsymbol{x}^{(i)}$

Algorithm 2: Forward-backward algorithm for synthesis

Input: $\boldsymbol{y} \in \mathbb{R}^M$, $\boldsymbol{a}^{(0)} \in \mathbb{R}^L$, σ and $\lambda^{(i)} \in (0, \infty)$
Output: $\boldsymbol{a}'$
do
  compute $\boldsymbol{u} = \boldsymbol{a}^{(i)} - \lambda^{(i)}\boldsymbol{\Psi}^\dagger\boldsymbol{\Phi}^\dagger(\boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{a}^{(i)} - \boldsymbol{y})/\sigma^2$
  update $\boldsymbol{a}^{(i+1)} = \operatorname{soft}_{\lambda^{(i)}\mu}(\boldsymbol{u})$
  $i = i + 1$
while *Stopping criterion is not reached*;
set $\boldsymbol{a}' = \boldsymbol{a}^{(i)}$

4 BAYESIAN UNCERTAINTY QUANTIFICATION: MAP ESTIMATION

The analysis and synthesis reconstruction models address inverse problems which are generally ill-conditioned or ill-posed (especially when the measurements are only observed partially or with limited resolution). Consequently, the corresponding estimators have significant intrinsic uncertainty that is very challenging to analyse and quantify. In Pereyra (2017) a general methodology was proposed to use MAP estimators to accurately approximate Bayesian credible regions for $p(\boldsymbol{x}|\boldsymbol{y})$. These credible regions indicate the regions of the parameter space where most of the posterior probability mass lies. A remarkable property of the approximation is that it only requires knowledge of $\boldsymbol{x}_{\text{map}}$ and therefore it can be computed very efficiently, even in very large-scale problems.

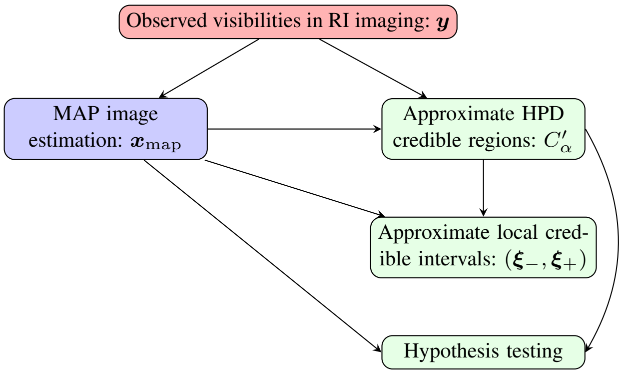


Figure 1. Our proposed uncertainty quantification procedure for RI imaging based on MAP estimation. The light green areas on the right show the types of uncertainty quantification developed. Firstly, an image is reconstructed by MAP estimation using convex optimisation techniques, which scale to big-data. Then, various forms of uncertainty quantification are performed. Global approximate Bayesian credible regions are computed. These are then used to compute local credible intervals (*cf.* error bars) corresponding to individual pixels and superpixels and to perform hypothesis testing of image structure to test whether a structure is physical or an artefact.

The diagram in Figure 1 shows the main components of our proposed uncertainty quantification methodology based on MAP estimation. As is shown, firstly, an image is reconstructed by MAP estimation. MAP estimation can be computed extremely rapidly and is therefore ideal for application to big-data. Then, various forms of uncertainty quantification are performed. Firstly, global approximate Bayesian credible regions are computed. These are then used to compute local credible intervals (*cf.* error bars) corresponding to individual pixels and superpixels. Finally, again using the global approximate Bayesian credible regions, hypothesis testing of image structure can be performed to test whether a structure is physical or an artefact. For consistency, we adopt the same notation as in the companion article (Cai et al. 2017a).

4.1 Approximate highest posterior density (HPD) credible regions

The first step in our uncertainty quantification methodology is to compute a credible region for $p(\boldsymbol{x}|\boldsymbol{y})$. A posterior credible region with credible level $100(1-\alpha)\%$ is a set $C_\alpha \in \mathbb{R}^N$ that satisfies

$$p(\boldsymbol{x} \in C_\alpha|\boldsymbol{y}) = \int_{\boldsymbol{x}\in\mathbb{R}^N} p(\boldsymbol{x}|\boldsymbol{y}) \mathbb{1}_{C_\alpha} \mathrm{d}\boldsymbol{x} = 1-\alpha, \tag{16}$$

where $\mathbb{1}_{C_\alpha}$ is the indicator function for C_α, defined by $\mathbb{1}_{C_\alpha}(\boldsymbol{u}) = 1$ if $\boldsymbol{u} \in C_\alpha$ and 0 otherwise. Many regions satisfy the above property. We focus on the HPD (Highest Posterior Density) region defined by

$$C_\alpha := \{\boldsymbol{x} : f(\boldsymbol{x}) + g(\boldsymbol{x}) \le \gamma_\alpha\}, \tag{17}$$

where the threshold γ_α which defines an isocontour or level-set of the log-posterior is set such that (16) holds, and we recall that $p(\boldsymbol{x}|\ \boldsymbol{y}) \propto \exp\{-f(\boldsymbol{x}) - g(\boldsymbol{x})\}$. This region is decision-theoretically optimal in the sense of minimum volume (Robert 2007).

Computing HPD credible regions in (17) is difficult because of the high-dimensional integral in (16). For RI models that are not too high dimensional, C_α can be computed efficiently by using proximal MCMC method as described in Cai et al. 2017a. However, this is not possible in big-data settings.

Here we use an approximation of C_α proposed recently in Pereyra (2017) for convex inverse problems solved by MAP estimation. The approximation is given by

$$C'_\alpha := \{\boldsymbol{x} : f(\boldsymbol{x}) + g(\boldsymbol{x}) \le \gamma'_\alpha\}, \tag{18}$$

where γ'_α is an approximation of the HPD threshold γ_α given by

$$\gamma'_\alpha = f(\boldsymbol{x}_{\text{map}}) + g(\boldsymbol{x}_{\text{map}}) + \tau_\alpha\sqrt{N} + N, \tag{19}$$

with universal constant $\tau_\alpha = \sqrt{16\log(3/\alpha)}$. Recall that N is the dimension of $\boldsymbol{x}$ and $100(1-\alpha)\%$ the credible level considered. After computing $\boldsymbol{x}_{\text{map}}$ by using modern convex optimisation algorithms, γ'_α can be calculated straightforwardly using (19), even in very high dimensions. The approximation given in (19) was motivated from recent results in information theory in terms of a probability concentration inequality (refer to Pereyra 2017 for more details).

For any $\alpha \in (4\mathrm{exp}(-N/3), 1)$, the error between γ'_α and γ_α is bounded by the following inequality

$$0 \le \gamma'_\alpha - \gamma_\alpha \le \eta_\alpha\sqrt{N} + N, \tag{20}$$

where $\eta_\alpha = \sqrt{16\log(3/\alpha)} + \sqrt{1/\alpha}$. Since the error $\gamma'_\alpha - \gamma_\alpha$ grows at most linearly with respect to N when N is large, the credible region C'_α associated with γ'_α is a stable approximation of C_α. Moreover, since $\gamma'_\alpha - \gamma_\alpha \ge 0$ the approximation is theoretically conservative in the sense that C'_α overestimates C_α. Precisely, in the analysis formulation, we first compute the reconstructed image $\boldsymbol{x}_{\text{map}}$ by using Algorithm 1, and then obtain an approximate HPD credible region

$$\bar{C}'^{,\text{map}}_\alpha := \{\boldsymbol{x} : \bar{f}(\boldsymbol{x}) + \bar{g}(\boldsymbol{x}) \le \bar{\gamma}'_\alpha\} \tag{21}$$

with

$$\bar{\gamma}'_\alpha = \bar{f}(\boldsymbol{x}_{\text{map}}) + \bar{g}(\boldsymbol{x}_{\text{map}}) + \tau_\alpha\sqrt{N} + N. \tag{22}$$

Similarly, in the synthesis setting we compute $\boldsymbol{a}_{\text{map}}$ via Algorithm 2, and then construct

$$\hat{C}'^{,\text{map}}_\alpha := \{\boldsymbol{\Psi}\boldsymbol{a} : \hat{f}(\boldsymbol{a}) + \hat{g}(\boldsymbol{a}) \le \hat{\gamma}'_\alpha\} \tag{23}$$

with

$$\hat{\gamma}'_\alpha = \hat{f}(\boldsymbol{a}_{\text{map}}) + \hat{g}(\boldsymbol{a}_{\text{map}}) + \tau_\alpha\sqrt{N} + N. \tag{24}$$

Note that $\bar{\gamma}'_\alpha$ and $\hat{\gamma}'_\alpha$ define the HPD credible regions implicitly.

The HPD credible regions can be used to quantify uncertainties in a variety of manners. In the reminder of this section we describe two such strategies.

4.2 Local credible intervals

The first strategy we propose is a novel approach to compute local credible intervals corresponding to pixels and superpixels, as a means for quantifying uncertainty spatially at different scales. This presents a new form of Bayesian uncertainty quantification tailored for image data and is easy to visualise and interpret. The method is based on the HPD credible regions discussed above and is applicable for any method for which HPD credible regions can be computed. Here we promote the MAP-based approach, based on the approximations (22) and (24), and benchmark our results against the MCMC approach Px-MALA, introduced in Cai et al. (2017a).

Let $\Omega = \cup_i\Omega_i$ be a partition of the image domain Ω into subsets or *superpixels* Ω_i such that $\Omega_i \cap \Omega_j = \emptyset, i \neq j$. The image domain can be partitioned at different scales, from a single pixel to larger scales involving blocks of several pixels. To index superpixels we define the index operator $\boldsymbol{\zeta}_{\Omega_i} = (\zeta_1, \cdots, \zeta_N) \in \mathbb{R}^N$ on Ω_i, which satisfies

$$\zeta_k = \begin{cases} 1, & \text{if } k \in \Omega_i, \\ 0, & \text{otherwise.} \end{cases} \tag{25}$$

To quantify the uncertainty associated with the region Ω_i we calculate the points ξ_{-,Ω_i} and ξ_{+,Ω_i} that saturate the HPD credible region $C'^{,\mathrm{map}}_{\alpha}$ from above and from below at Ω_i, given by

$$\xi_{-,\Omega_i} = \min_{\xi} \left\{ \xi | f(\boldsymbol{x}_{i,\xi}) + g(\boldsymbol{x}_{i,\xi}) \leq \gamma'_\alpha, \forall \xi \in [0, +\infty) \right\}, \tag{26}$$

$$\xi_{+,\Omega_i} = \max_{\xi} \left\{ \xi | f(\boldsymbol{x}_{i,\xi}) + g(\boldsymbol{x}_{i,\xi}) \leq \gamma'_\alpha, \forall \xi \in [0, +\infty) \right\}, \tag{27}$$

where $\boldsymbol{x}_{i,\xi} = \boldsymbol{x}^*(\mathbf{I} - \boldsymbol{\zeta}_{\Omega_i}) + \xi \boldsymbol{\zeta}_{\Omega_i}$ represents a point estimator generated by replacing the intensity of $\boldsymbol{x}^*$ in Ω_i by ξ. We recall that γ'_α is the threshold or isocontour level defining $C'^{,\mathrm{map}}_{\alpha}$. We then construct the interval $(\xi_{-,\Omega_i}, \xi_{+,\Omega_i})$ that represents the range of intensity values ξ of Ω_i for which $\boldsymbol{x}_{i,\xi} \in C'^{,\mathrm{map}}_{\alpha}$.

Finally, for visualisation, we gather all the lower and upper bounds $\xi_{-,\Omega_i}, \xi_{+,\Omega_i}, \forall i$, into the following two images:

$$\boldsymbol{\xi}_- = \sum_i \xi_{-,\Omega_i} \boldsymbol{\zeta}_{\Omega_i}, \quad \boldsymbol{\xi}_+ = \sum_i \xi_{+,\Omega_i} \boldsymbol{\zeta}_{\Omega_i}. \tag{28}$$

We typically consider the difference image $(\boldsymbol{\xi}_+ - \boldsymbol{\xi}_-)$ that shows the length of the local credible intervals (*cf.* error bars). These images can be constructed at different scales to analyse structure of different sizes. In our experiments, as examples, we consider superpixels of sizes 10×10, 20×20, and 30×30 pixels.

To conclude, notice that visualising uncertainty in high dimensional problems is fundamentally difficult. For example, even the simple case of N-dimensional Gaussian models involves covariance matrices of size $N \times N$; the models considered here are significantly more complex. As a result, uncertainty information could potentially structure along directions of the parameter space that the visual uncertainty plots described above fail to capture. However, we believe that correlations in images are predominantly local, albeit at potentially different scales. What our analyses seek to capture and visually display are precisely these local correlations at superpixel scales of different levels.

4.3 Hypothesis testing of image structure

In a manner akin to the companion article Cai et al. (2017a), we use *knock-out* posterior tests to assess specific areas or structures of interest in the reconstructed images. These tests proceed by constructing a surrogate test image $\boldsymbol{x}^{*,\mathrm{sgt}}$ by carefully replacing the structure of interest in an point estimator $\boldsymbol{x}^*$ (or $\boldsymbol{\Psi}\boldsymbol{a}^*$) with background information. If removing the structure has pushed $\boldsymbol{x}^{*,\mathrm{sgt}}$ outside of the HPD credible region (*i.e.* $\boldsymbol{x}^{*,\mathrm{sgt}} \notin C'^{,\mathrm{map}}_{\alpha}$), this indicates that the data strongly supports the structure under consideration. Conversely, if $\boldsymbol{x}^{*,\mathrm{sgt}}$ remains inside of the HPD credible region (i.e. $\boldsymbol{x}^{*,\mathrm{sgt}} \in C'^{,\mathrm{map}}_{\alpha}$), then the likelihood is insensitive to the modification, indicating lack of strong evidence for the scrutinised structure.

Algorithmically, a surrogate $\boldsymbol{x}^{*,\mathrm{sgt}}$ for a test area $\Omega_D \subset \Omega$ is generated by performing segmentation-inpainting of $\boldsymbol{x}^*$, for example by applying a wavelet filter $\boldsymbol{\Lambda}$ iteratively by using

$$\boldsymbol{x}^{(m+1),\mathrm{sgt}} = \boldsymbol{x}^* \mathbb{1}_{\Omega - \Omega_D} + \boldsymbol{\Lambda}^\dagger \mathrm{soft}_{\lambda_{\mathrm{thd}}}(\boldsymbol{\Lambda}\boldsymbol{x}^{(m),\mathrm{sgt}}) \mathbb{1}_{\Omega_D}, \tag{29}$$

with $\boldsymbol{x}^{(0),\mathrm{sgt}} = \boldsymbol{x}^*$ or $\boldsymbol{x}^{(0),\mathrm{sgt}} = \boldsymbol{\Psi}\boldsymbol{a}^*$ for the synthesis formulation (usually 100 iterations are sufficient for convergence). To determine if $\boldsymbol{x}^{*,\mathrm{sgt}} \in C'^{,\mathrm{map}}_{\alpha}$, it suffices to check if

$$f(\boldsymbol{x}^{*,\mathrm{sgt}}) + g(\boldsymbol{x}^{*,\mathrm{sgt}}) \leq \gamma'_\alpha. \tag{30}$$

In addition to the approach presented above to assess the existence of specific areas or structures of interest, we also propose the following approach to focus on assessing sub-structure within areas of interest. Briefly speaking, we create surrogate test images with the sub-structure in question effectively removed by smoothing the corresponding region. Algorithmically, a surrogate $\boldsymbol{x}^{*,\mathrm{sgt}}$ for a test area $\Omega_D \subset \Omega$ is then generated by

$$\boldsymbol{x}^{*,\mathrm{sgt}} = \boldsymbol{x}^* \mathbb{1}_{\Omega - \Omega_D} + (\mathbf{S}\boldsymbol{x}^*) \mathbb{1}_{\Omega_D}, \tag{31}$$

where $\mathbf{S}$ is a smoothing operator applied to remove sub-structure within the test area Ω_D.

5 EXPERIMENTAL RESULTS

We now investigate the performance of the proposed uncertainty quantification methodology for the three strategies discussed in Section 4. We also report a detailed comparison with the proximal MCMC method Px-MALA, which is one of the MCMC methods introduced in the companion article (Cai et al. 2017a) and that can also support sparsity-promoting priors. Px-MALA produces (asymptotically) exact inferences and therefore we use it here as an accurate benchmark for the methods proposed in this article.

5.1 Simulations

In a manner akin to Cai et al. (2017a), we perform our experiments with the following four RI images: M31 galaxy (size 256×256), Cygnus A galaxy (size 256×512), W28 supernova remnant (size 256×256), and 3C288 (size 256×256). These images are depicted in Figure 2 (a) and Figure 3 (a). Radio interferometric observations are simulated for these ground truth images in a similar manner as in Cai et al. (2017a).

The numerical experiments performed in this article for MAP estimation were run on a Macbook laptop with an i7 Intel CPU and memory of 16 GB, running MATLAB R2015b. The Px-MALA algorithm used as a benchmark is significantly more computationally expensive and required a high-performance workstation (see Cai et al. 2017a). For further details about the experiment setup and the implementation of Px-MALA please see Cai et al. (2017a).

Regarding the models used for the experiments, the ℓ_1 regularisation parameter μ in the analysis and synthesis models is set to 10^4 and the dictionary $\boldsymbol{\Psi}$ in the analysis and synthesis models is set to Daubechies 8 wavelets. In Algorithms 1 and 2, we use $\lambda^{(i)} = 0.5$, with stopping criteria set by a maximum iteration number of 500 or relative difference between solutions of 10^{-4}. In formulas (22) and (24), the range of values for α is $[0.01, 0.99]$. In particular, credible regions and intervals are reported at $\alpha = 0.01$, corresponding to the 99% credible level. The maximum number of iterations for segmented-inpainting in (29) is set to 200.

5.2 Image reconstruction

As the first step in our analysis we perform Bayesian image reconstruction for the four images considered. Precisely, for each image we compute two Bayesian estimators, the MAP estimator computed by convex optimisation and the sample mean estimator

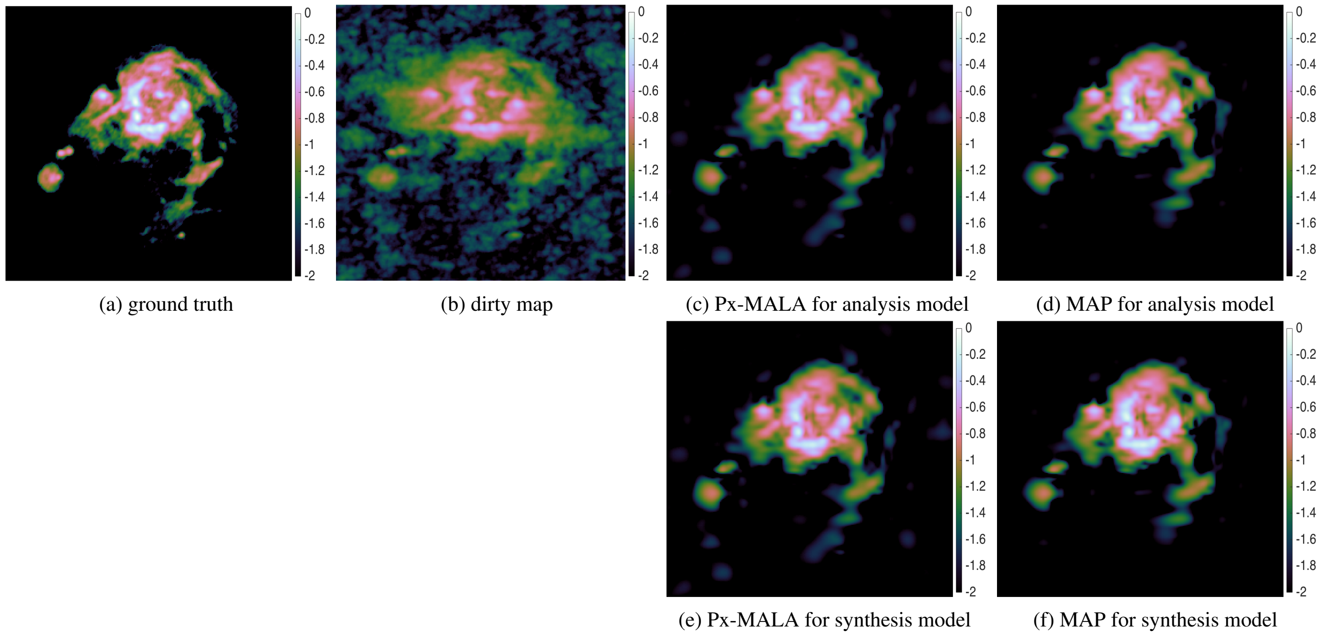


Figure 2. Image reconstructions for M31 (size 256×256). All images are shown in $\mathtt{log}_{10}$ scale (*i.e.* the numeric labels on the colour bar are the logarithms of the image intensity).. (a): ground truth; (b): dirty image (reconstructed by inverse Fourier transform); (c) and (d): point estimators for the analysis model (3) computed by Px-MALA and MAP estimation, respectively; (e) and (f): the same as (c) and (d) but for the synthesis model (4). In particular, the point estimators of Px-MALA are the sample mean. Clearly, consistent results between Px-MALA and MAP estimation and between the analysis and synthesis models are obtained.

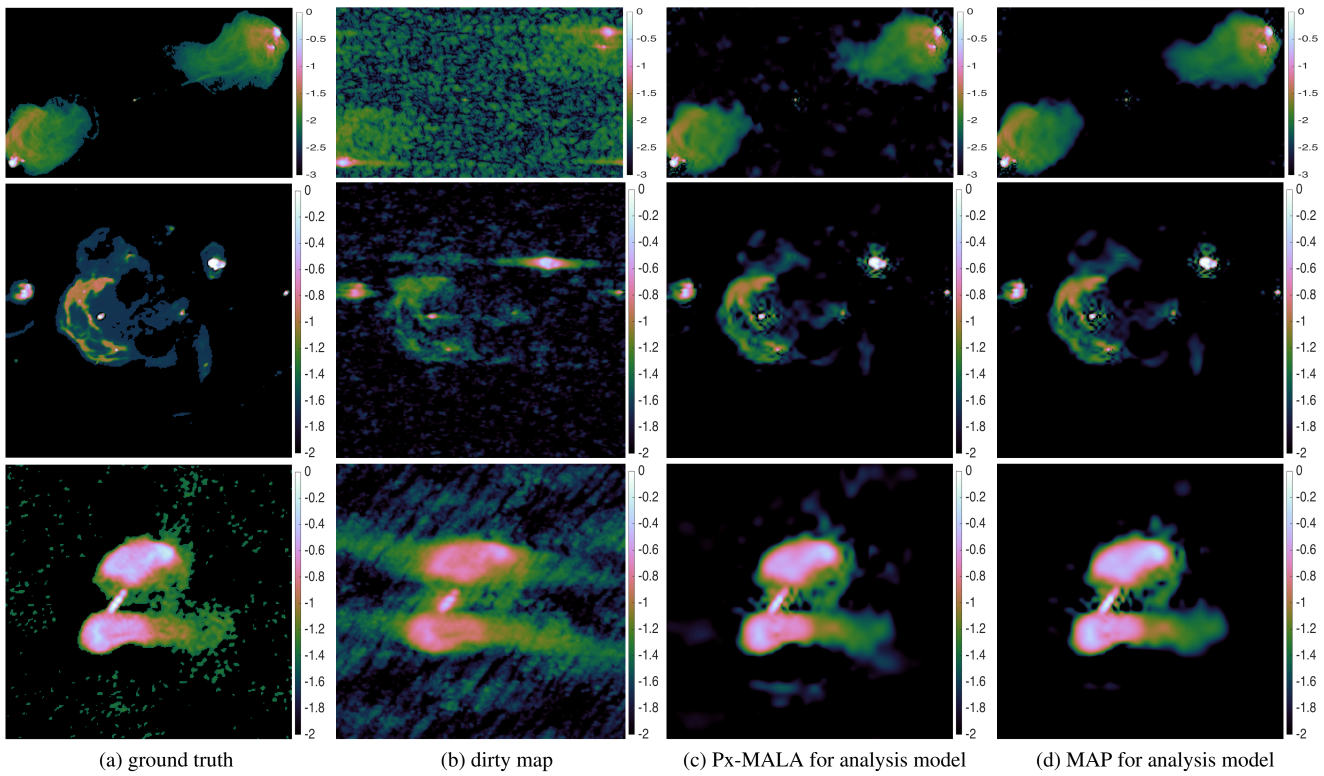


Figure 3. Image reconstructions for Cygnus A (size 256×512), W28 (size 256×256), and 3C288 (size 256×256) radio galaxies (first to third rows). All images are shown in $\mathtt{log}_{10}$ scale. First column: (a) ground truth. Second to forth columns: (b) dirty images; (c) and (d) point estimators for the analysis model (3) computed by Px-MALA and MAP estimation, respectively. Clearly, consistent results between Px-MALA and MAP estimation are obtained.

Table 1. CPU time in minutes for the proximal MCMC method Px-MALA (generating full posterior samples) and MAP-based methods (computing a point estimator), for the analysis and synthesis models and for test images of M31, Cygnus A, W28 and 3C288. MAP estimation is approximately 10^5 times faster than Px-MALA and can be scaled to big-data.

Images	Methods	CPU time (min) Analysis	Synthesis
M31 (Fig. 2)	Px-MALA	1307	944
	MAP	.03	.02
Cygnus A (Fig. 3)	Px-MALA	2274	1762
	MAP	.07	.04
W28 (Fig. 3)	Px-MALA	1122	879
	MAP	.06	.04
3C288 (Fig. 3)	Px-MALA	1144	881
	MAP	.03	.02

computed with Px-MALA. For completeness, we consider both the analysis and the synthesis models (3) and (4).

The Bayesian estimators related to the analysis model are shown in Figures 2 and 3. Observe that both estimators produce similar, excellent reconstruction results. For comparison, dirty maps (reconstructed by applying the inverse Fourier transform directly to the visibilities) of the test images are shown in Figure 2 (b) and Figure 3 (b). As expected, the results of the analysis and synthesis models (3) and (4) under an orthogonal basis $\boldsymbol{\Psi}$ are nearly undistinguishable[6] (see results for M31 in Figure 2; to avoid redundancy the results for the other images are not reported here). For this reason, in the reminder of this article only the results for the analysis model are presented.

We emphasise again that MAP estimators computed by convex optimisation are significantly faster to compute than the estimators that require MCMC methods. In particular, in our experiments there is a gain of order 10^5 in terms of computation time (see Table 1 for the computation time comparisons with Px-MALA). Furthermore, MAP estimation based on convex optimisation supports algorithmic structures that can be highly distributed (*e.g.* Carrillo et al. 2014; Onose et al. 2016) to further assist in scaling to big-data. MCMC algorithms cannot typically be distributed to such a high degree. We have not yet considered distributed MAP algorithms here; our MAP-based methods therefore provide additional performance improvements over MCMC beyond the already dramatic improvements shown in Table 1.

5.3 Approximate HPD credible regions

We compute the HPD credible regions for the four images considered. Precisely, we use formulas (22) and (24) to approximate the threshold or isocontour value γ'_α defining the HPD regions for the analysis and synthesis models (recall that these are highly efficient approximations derived from the MAP estimates $\boldsymbol{x}_{\rm map}$ and $\boldsymbol{a}_{\rm map}$). Figure 4 shows the threshold values obtained for each image and model, for $\alpha \in [0.01, 0.99]$; observe again that the results of the analysis and synthesis models are consistent with each other, as expected.

[6] Note that, when $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} = \mathsf{I}$, as considered here, the analysis and synthesis models are identical. However, when $\boldsymbol{\Psi}^\dagger\boldsymbol{\Psi} \neq \mathsf{I}$, they are very different and we expect different reconstructed images.

To assess the approximation error involved in using the approximations (22) and (24), we also computed the exact HPD threshold values by using the Px-MALA MCMC algorithm (*cf.* Cai et al. 2017a, Figure 6). Recall than Px-MALA is several orders of magnitude more computationally expensive than MAP estimation (see Table 1). This comparison revealed approximation errors of between 1% and 5% over all cases, which is in close agreement with the results reported in Pereyra (2017). These experiments confirm that the MAP-based approximations (22) and (24) deliver accurate estimates of the HPD credible regions with a dramatically lower computational cost.

5.4 Approximate local credible intervals

We use the approximate HPD regions to calculate local credible intervals for image superpixels. Precisely, Figures 5–8 report the length of local credible intervals for the four test images for superpixel grid sizes of 10×10, 20×20, and 30×30 pixels, computed w.r.t. the analysis model (the results for the synthesis model are very similar). For comparison, Figures 5–8 also show the exact local credible estimates obtained by using the Px-MALA MCMC algorithm, which does not rely on the approximations (22) and (24).

We conclude the main observations as follows. Firstly, the results obtained with both approaches are extremely consistent with each other, indicating that the approximate credible intervals derived from the MAP estimation are very accurate. Secondly, the length of the approximate local credible intervals computed by MAP estimation are theoretically conservative and can be seen to slightly overestimate the lengths computed by MCMC sampling, and so are trustworthy. Thirdly, note that (i) coarser scales have shorter credible intervals than narrower scales, and (ii) superpixels at object boundaries generally have longer credible intervals than superpixels in homogenous regions. These two observations are related to the fact that narrow scales are mainly sensitive to high spatial frequency information such as fine details and object boundaries that are difficult to accurately estimate, whereas coarser scales are also sensitive to lower frequencies and larger structures that are easier to estimate. More precisely, these two observations are a direct consequence of the fact that the sampling profile associated with the measurement operator $\boldsymbol{\Phi}$ mainly covers low frequencies and has very few high-frequency measurements (see Cai et al. 2017a, Figure 2). As a result, the likelihood $p(\boldsymbol{y}|\boldsymbol{x})$ has significantly less information about high-frequency image components, and this leads to higher uncertainty (*i.e.*, longer credible intervals) at fine scales, sharp details, and object boundaries.

5.5 Hypothesis testing of image structure

We conclude our experimental results by demonstrating our methodology for testing structure in reconstructed images. We consider the same images and structures of interest as in Cai et al. (2017a), shown in the yellow rectangular areas in the first column of Figure 9. All of these structures are physical (*i.e.* present in the ground truth images), except for structure 2 in 3C288 which is a reconstruction artefact.

Recall that the methodology proceeds as follows. First, we construct a carefully designed surrogate image $\boldsymbol{x}^{*,\rm sgt}$ by modifying the MAP estimator $\boldsymbol{x}_{\rm map}$ to remove the structure of interest via segmentation-inpaiting, computed using formula (29) (notice that this modification produces a surrogate that is in agreement with the prior distribution). Each structure is assessed individually. Second,

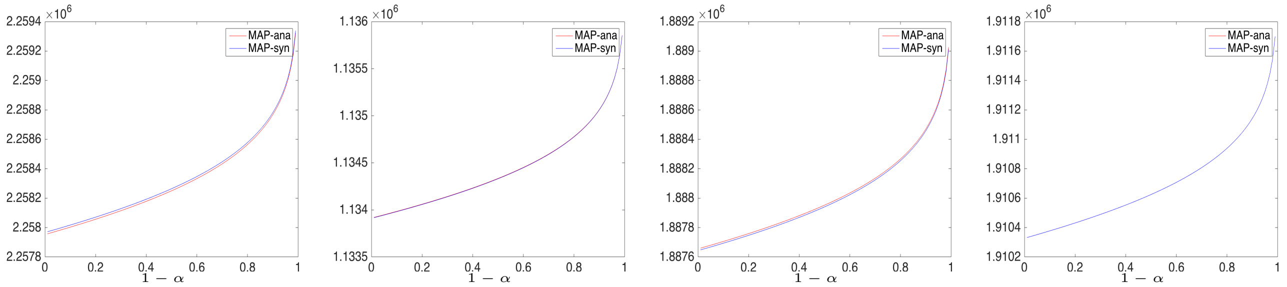


Figure 4. HPD credible region isocontour levels $\bar{\gamma}'_\alpha$ and $\hat{\gamma}'_\alpha$ computed using MAP-based methods, for test images (a) M31, (b) Cygnus A, (c) W28, and (d) 3C288. In particular, MAP-ana (*resp.* MAP-syn) represents the results by MAP estimation for the analysis (*resp.* synthesis) model. Note that the red line in plot (d) is overlaid by the blue line and thus may not be visible, due to the high degree of similarity between the two results. In all cases the results of the analysis and synthesis models are in close agreement.

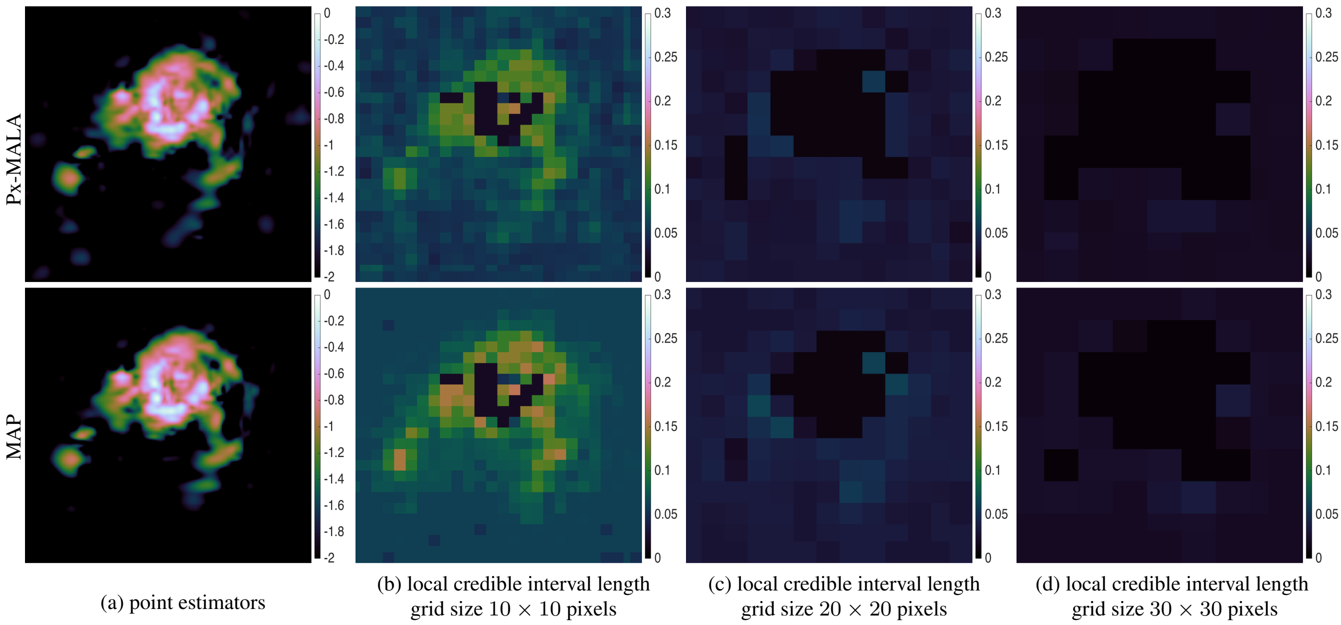


Figure 5. Length of local credible intervals (99% credible level), *cf.* error bars, computed for M31 for the analysis model (3). First column: (a) point estimators. Second to fourth columns: (b)–(d) local credible intervals at grid sizes of 10×10, 20×20, and 30×30 pixels, respectively. First row gives exact inferences computed with the MCMC method Px-MALA (Cai et al. 2017a). Second row gives MAP-based approximate inferences computed by convex optimisation. Clearly, MAP-based approximations provide estimates of the length of local credible intervals (*cf.* error bars) that are extremely consistent with the ones obtained by Px-MALA, while the MAP estimates can be computed several orders of magnitude more rapidly (Table 1). Moreover, the length of the approximate credible intervals computed by the MAP-based approach are theoretically conservative and can be seen to slightly overestimate the lengths computed by MCMC sampling.

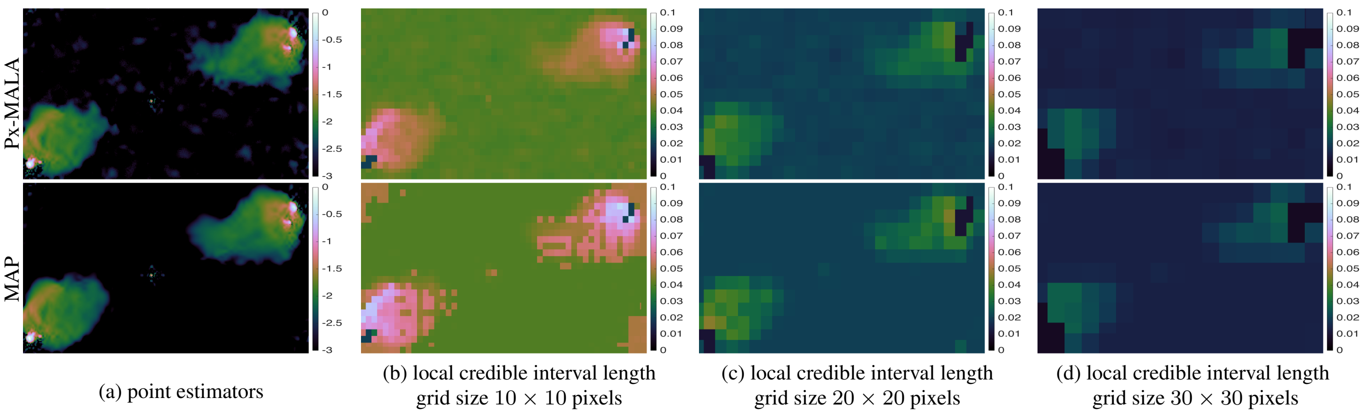
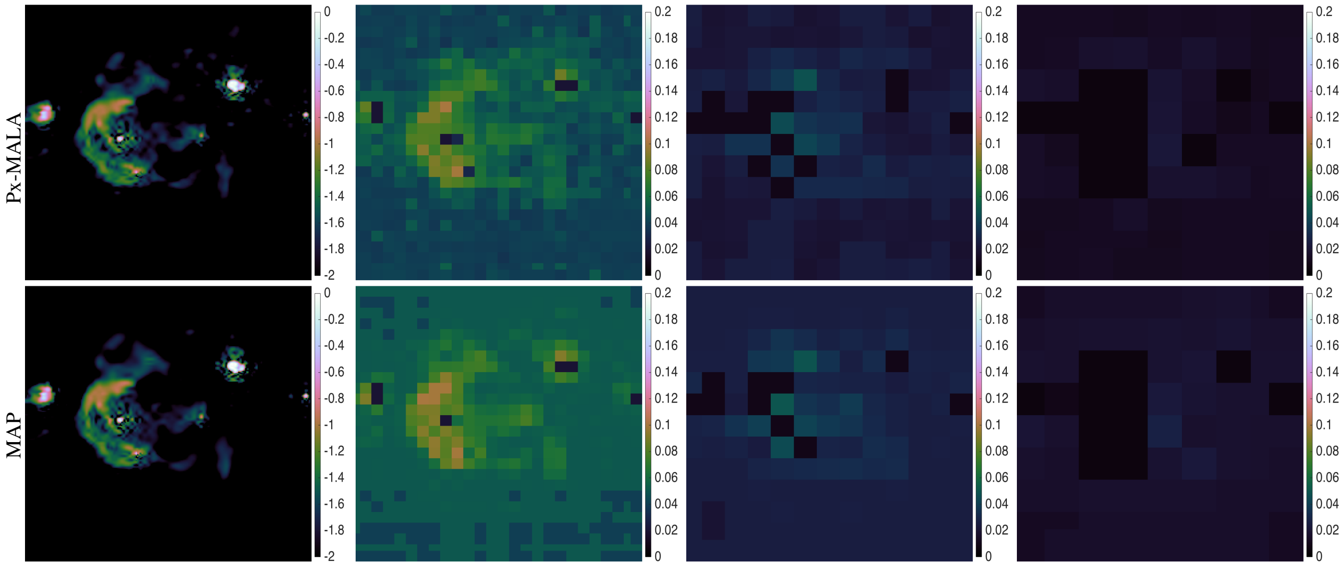
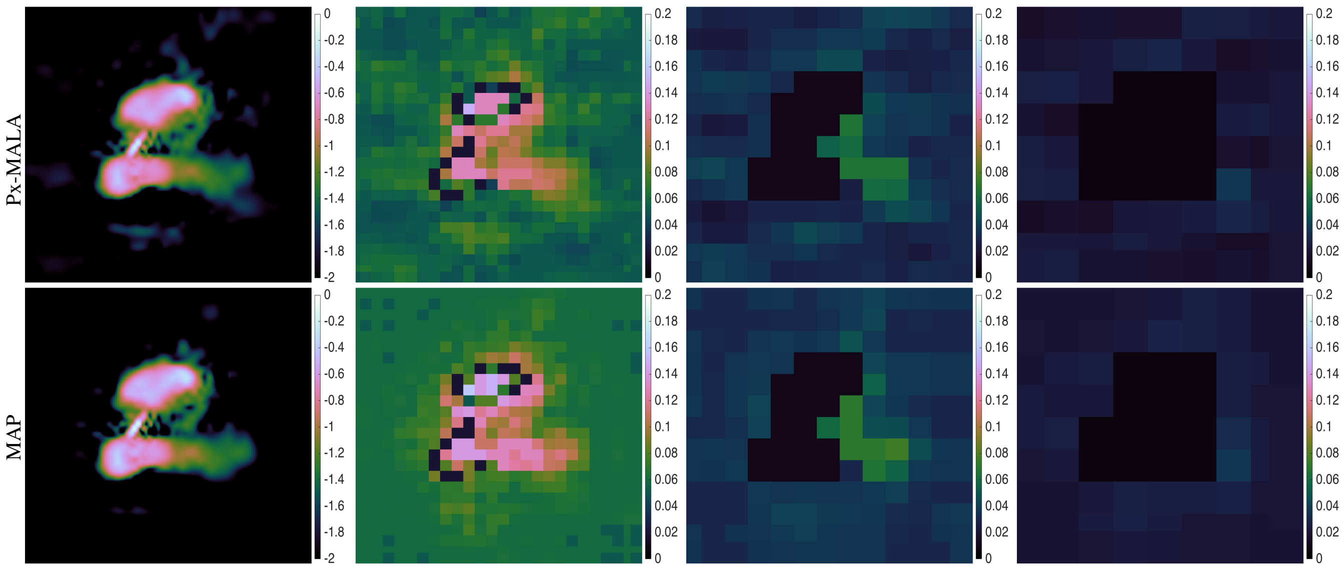


Figure 6. Same as Figure 5 but for Cygnus A.



(a) point estimators | (b) local credible interval length grid size 10 × 10 pixels | (c) local credible interval length grid size 20 × 20 pixels | (d) local credible interval length grid size 30 × 30 pixels

Figure 7. Same as Figure 5 but for W28.



(a) point estimators | (b) local credible interval length grid size 10 × 10 pixels | (c) local credible interval length grid size 20 × 20 pixels | (d) local credible interval length grid size 30 × 30 pixels

Figure 8. Same as Figure 5 but for 3C288.

we check if $\boldsymbol{x}^{*,\mathrm{sgt}} \notin C'^{,\mathrm{map}}_{\alpha}$ (*i.e.* if $f(\boldsymbol{x}^{*,\mathrm{sgt}}) + g(\boldsymbol{x}^{*,\mathrm{sgt}}) > \gamma'_{\alpha}$) to determine whether the surrogate falls within the conservative HPD credible region or not. The resulting surrogate images are displayed in the second column of Figure 9. If the fact of removing the structure from $\boldsymbol{x}_{\mathrm{map}}$, which is at the centre of $C'^{,\mathrm{map}}_{\alpha}$, produces a surrogate that is outside $C'^{,\mathrm{map}}_{\alpha}$, this indicates that the likelihood is in clear disagreement with that modification. In that case we conclude that there exists significant evidence in the observed data in favour of the structure considered. Otherwise, we conclude that we fail to establish that there is significant evidence in favour of that structure. We emphasise at this point that conclusions are generally not highly sensitive to the exact value of α; here we report results for $\alpha = 0.01$ related to a 99% credible level.

The results of these tests are shown in Table 2. For comparison, we also include the results obtained with the reference method Px-MALA (Cai et al. 2017a). Again, the two methods produce excellent results that are consistent with each other. From Table 2, we observe that the methods have correctly classified the three main physical structures of M31, W28, and 3C288, and correctly identified the minor structure of 3C288 as a potential reconstruction artefact. Moreover, the methods have found that it is not possible to make a strong statistical statement about the small physical structure in image Cygnus A, which is difficult because it is only a few pixels in size, isolated, and significantly weaker in intensity than the other structures in the image.

To test the performance of hypothesis testing in terms of as-

Table 2. Hypothesis test results for test structures shown in Figure 9 for M31, Cygnus A, W28, and 3C288. Note that γ_α represents the isocontour defining the HPD credible region at credible level $(1-\alpha)$, where here $\alpha = 0.01$, $\boldsymbol{x}^{*,\rm sgt}$ represents the surrogate generated from point estimator $\boldsymbol{x}^*$ (in particular, for Px-MALA $\boldsymbol{x}^*$ is the sample mean of the MCMC samples), and $(f+g)(\cdot)$ represents the objective function; symbols with labels $^-$ and $\hat{}$ are related to the analysis model (3) and the synthesis model (4), respectively. Symbol ✗ indicates that the test area is artificial (and no strong statistical statement can be made as to the area), while ✓ indicates that the test area is physical. All values are in units 10^6. Clearly, both Px-MALA and MAP estimation give convincing and consistent hypothesis test results. Note that MAP estimation is dramatically more computationally efficient that Px-MALA (Table 1).

Images	Test areas	Ground truth	Method	$(\bar{f}+\bar{g})(\bar{\boldsymbol{x}}^{*,\rm sgt})$	Isocontour $\bar{\gamma}_{0.01}$	$(\hat{f}+\hat{g})(\boldsymbol{\Psi}^\dagger\hat{\boldsymbol{x}}^{*,\rm sgt})$	Isocontour $\hat{\gamma}_{0.01}$	Hypothesis test
M31 (Fig. 9)	1	✓	Px-MALA	**2.44**	2.34	**2.43**	2.34	✓
			MAP	**2.29**	2.26	**2.29**	2.26	✓
Cygnus A (Fig. 9)	1	✓	Px-MALA	1.17	**1.26**	1.18	**1.27**	✗
			MAP	1.02	**1.14**	1.02	**1.14**	✗
W28 (Fig. 9)	1	✓	Px-MALA	**3.38**	1.84	**3.37**	1.85	✓
			MAP	**3.47**	1.89	**3.47**	1.89	✓
3C288 (Fig. 9)	1	✓	Px-MALA	**3.27**	2.02	**3.25**	2.01	✓
			MAP	**3.11**	1.91	**3.11**	1.91	✓
	2	✗	Px-MALA	1.971	**2.027**	1.954	**2.010**	✗
			MAP	1.844	**1.912**	1.844	**1.912**	✗

sessing sub-structure within areas of interest, we consider sub-structure in an area in M31 (see Figure 10). We find that the surrogate test image shown in Figure 10 (b) falls outside of the HPD credible region (the objective of the surrogate is 2.38×10^6, which is larger than the HPD isocontour of $\bar{\gamma}_{0.01} = 2.26 \times 10^6$) according to the analysis model (the hypothesis testing result regarding the synthesis model is the same). Therefore, the sub-structure shown in the specified area in Figure 10 (a) is correctly classified as physical at a high credible level.

Before closing this section, we emphasise again that the methods presented in this article deliver a variety of forms of uncertainty quantification with a very low computational cost. While these new forms of uncertainty quantification can also be achieved by using state-of-the-art proximal MCMC methods, such as Px-MALA and MYULA, as presented in the companion article Cai et al. (2017a), MCMC techniques cannot scale to massive data sizes. Nevertheless, they are useful for medium-scale problems and provide accurate benchmarks for the highly efficient methods presented herein, which will scale very well to the emerging big-data era of radio astronomy.

6 CONCLUSIONS

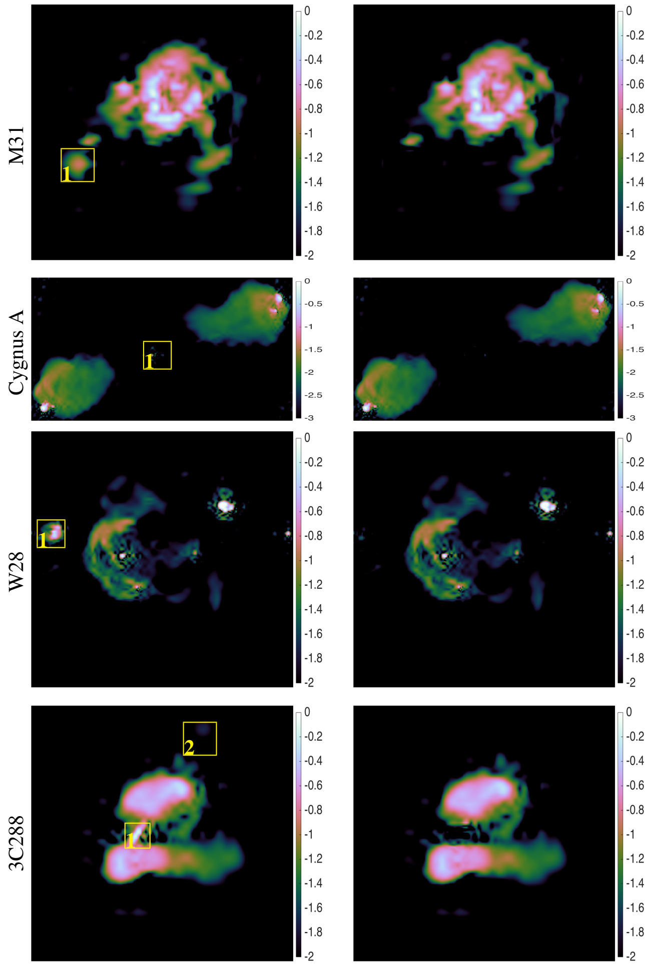
Uncertainty quantification is an important missing component in RI imaging that will only become increasingly important as the big-data era of radio interferometry emerges. No existing RI imaging techniques that are used in practice (*e.g.* CLEAN, MEM or CS approaches) provide uncertainty quantification. In this article, as an alternative to MCMC methods, such as Px-MALA and MYULA that were presented in Cai et al. (2017a), we present new uncertainty quantification methods MAP estimation by convex optimisation. The proposed uncertainty quantification methods exhibit extremely fast computation speeds and allow uncertainty quantification to be performed practically and in a manner that will scale to the emerging big-data era of RI imaging.

Our proposed methods, which inherit the advantages of convex optimisation methods, are much more efficient than proximal MCMC methods that explore the entire posterior distribution of the image. Note, however, that the methods proposed here give an approximation of HPD credible regions and, consequently, the additional forms of uncertainty quantification that are built on the approximate HPD credible regions are also approximate. Nevertheless, we show these approximations are very accurate. Moreover, the approximations are conservative so that uncertainties are not underestimated. In contrast, proximal MCMC methods can theoretically provide HPD credible regions and other forms of uncertainty quantification that are more accurate. Therefore, the proposed fast MAP-based methods and the proximal MCMC methods complement each other, rather than being mutually exclusive. We anticipate that when it comes to the big-data era, we will use predominantly fast uncertainty quantification methods such as those based on MAP estimation, and reserve MCMC methods for benchmarking and detailed comparison.

A variety of forms of uncertainty quantification for MAP estimation were constructed, including HPD credible regions, local credible intervals (*cf.* error bars) for individual pixels and superpixels, and tests for image structure. Our methods were evaluated on four test images that are representative in RI imaging. These experiments demonstrated that our MAP-based methods exhibit excellent performance and can reconstruct images with sharp detail. Moreover, they simultaneously underpin highly accurate approximate techniques to quantify uncertainties. In terms of computation time, MAP techniques were found to be approximately 10^5 times faster than state-of-the-art proximal MCMC methods, even when MAP estimation is run on a standard laptop and proximal MCMC methods on a high-performance workstation. Moreover, they lead to algorithmic structures that can be highly distributed and parallelised.

In the near future, we plan to apply the uncertainty quantification techniques presented in this article to RI observations acquired by a variety of different telescopes and to make the methods publicly available. The methods will be implemented in the existing PURIFY[7] package for RI imaging. Furthermore, novel algorithms will be developed to implement our methods with improved computational efficiency and to highly distribute and parallelise computations and data. We will also investigate optimal techniques for setting the regularisation parameter in a hierarchical Bayesian framework, applying the strategies developed by Pereyra et al. (2015).

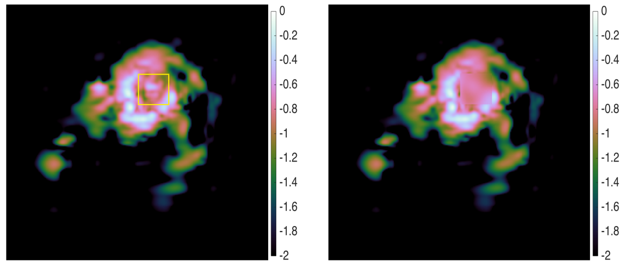
[7] https://github.com/basp-group/purify



(a) MAP point estimators (b) inpainted surrogate

Figure 9. Hypothesis testing of image structure for M31, Cygnus A, W28, and 3C288. The five structures depicted in yellow are considered, all of which are physical (*i.e.* present in the ground truth images), except for structure 2 in 3C288, which is a reconstruction artefact. First column (a): point estimators obtained by MAP estimation for the analysis model (3) (shown in $\log_{10}$ scale). Second column (b): segmented-inpainted surrogate test images with information in the yellow rectangular areas removed and replaced by inpainted background (shown in $\log_{10}$ scale). Hypothesis testing is then performed to test whether the structure considered is physical by checking whether the surrogate test images shown in (b) fall outside of the HPD credible regions. Results of these hypothesis tests are specified in Table 2. Note that for the case shown in the last row the structures within areas 1 and 2 are tested independently.

It is our hope that uncertainty quantification, *e.g.* in the form of recovering error bars (Bayesian credible intervals) and hypothesis testing of image structure and sub-structure, will become an important standard component in RI imaging for statistically principled and robust scientific inquiry. For the first time, we propose techniques for the practical quantification of uncertainties in RI imaging. These techniques can be applied not only to observations made by existing telescopes but also to the emerging big-data era of radio astronomy.



(a) MAP point estimator (b) smoothed surrogate

Figure 10. Hypothesis testing of image sub-structure for M31 (both images are shown in $\log_{10}$ scale). The area depicted in yellow is considered, where the sub-structure presented in it is physical in the ground truth image. First column (a): point estimator obtained by MAP estimation for the analysis model (3). Second column (b): smoothed surrogate test image with information in the yellow rectangular area smoothed (a MATLAB built-in function `imgaussfilt` using Gaussian filtering with standard deviation 6 is applied). Hypothesis testing is then performed to test whether the sub-structure in the area considered is physical by checking whether the surrogate test image shown in (b) falls outside of the HPD credible regions. The null hypothesis is rejected and the sub-structure of interested is correctly classified as physical and not a reconstruction artefact.

ACKNOWLEDGEMENTS

This work is supported by the UK Engineering and Physical Sciences Research Council (EPSRC) by grant EP/M011089/1, and Science and Technology Facilities Council (STFC) ST/M00113X/1. We also thank the editor and the anonymous reviewer for their constructive comments, which have significantly improved this manuscript.

APPENDIX A: CONVEX OPTIMISATION METHODS FOR MAP ESTIMATION

Forward-backward splitting algorithms solve optimisation problems of the form

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\operatorname{argmin}}(f+g)(\boldsymbol{x}), \tag{A1}$$

by using a splitting of $(f+g)(\boldsymbol{x})$. We consider the setting where $f \notin \mathcal{C}^1$ is proper, convex and lower semi-continuous (l.s.c.) and $g \in \mathcal{C}^1$ is l.s.c. convex and $\beta_{\rm Lip}$-Lipchitz differentiable, *i.e.*,

$$\|\nabla g(\hat{\boldsymbol{z}}) - \nabla g(\bar{\boldsymbol{z}})\| \le \beta_{\rm Lip}\|\hat{\boldsymbol{z}} - \bar{\boldsymbol{z}}\|, \ \ \forall(\hat{\boldsymbol{z}}, \bar{\boldsymbol{z}}) \in \mathbb{C}^N \times \mathbb{C}^N. \tag{A2}$$

Precisely, forward-backward algorithms solve (A1) by using the iteration

$$\boldsymbol{x}^{(i+1)} = \operatorname{prox}_{\lambda^{(i)} f}(\boldsymbol{x}^{(i)} - \lambda^{(i)}\nabla g(\boldsymbol{x}^{(i)})), \tag{A3}$$

where $\lambda^{(i)}$ is the step size in a suitable bounded interval (see, *e.g.*, Combettes & Pesquet 2010). The *proximity operator* of λf is defined as (Moreau 1965)

$$\operatorname{prox}_{\lambda f}(\boldsymbol{z}) \equiv \underset{\boldsymbol{u}\in\mathbb{R}^N}{\operatorname{argmin}}\left\{f(\boldsymbol{u}) + \|\boldsymbol{u} - \boldsymbol{z}\|^2/2\lambda\right\}. \tag{A4}$$

It is worth mentioning that when f is associated with the ℓ_1 norm, then computing (A4) goes to the so-called pointwise soft-thresholding of $\boldsymbol{z}$, *i.e.*, $\operatorname{soft}_\lambda(\boldsymbol{z}) = \big(\operatorname{soft}_\lambda(z_1), \operatorname{soft}_\lambda(z_2), \cdots\big)$ defined by

$$\operatorname{soft}_\lambda(z_j) = \begin{cases} z_j(|z_j| - \lambda)/|z_j| & \text{if } |z_j| > \lambda, \\ 0 & \text{otherwise,} \end{cases} \tag{A5}$$

for every component z_j.

There are several refinements of (A3) with better convergence properties. For example, using relaxation leads to the iteration

$$\boldsymbol{x}^{(i+1)} = (1-\beta^{(i)})\boldsymbol{x}^{(i)} + \beta^{(i)}\tilde{\boldsymbol{x}}^{(i+1)}, \tag{A6}$$

where $\tilde{\boldsymbol{x}}^{(i+1)}$ is computed by (A3), $\beta^{(i)}$ is a sequence of relaxation parameters, $\lambda^{(i)} \in (\epsilon, 2/\beta_{\rm Lip} - \epsilon)$, $\beta^{(i)} \in (\epsilon, 1)$, and $\epsilon \in (0, \min\{1, 1/\beta_{\rm Lip}\})$ (Combettes & Wajs 2005); or with $\lambda^{(i)} = 1/\beta_{\rm Lip}$, $\beta^{(i)} \in (\epsilon, 3/2 - \epsilon)$, and $\epsilon \in (0, 3/4)$ (Bauschke & Combettes 2011). Furthermore, algorithmic structures that allow computations to be highly distributed and parallelised (*e.g.* Carrillo et al. 2014; Onose et al. 2016) and computed in an online manner (Cai et al. 2017b) can also be developed to assist in scaling to big-data.

REFERENCES

Ables J. G., 1974, A&AS, 15, 383

Bauschke H. H., Combettes P. L., 2011, Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer-Verlag

Bhatnagar S., Corwnell T. J., 2004, A&A, 426, 747

Cai X., Pratley L., McEwen J. D., 2017b, preprint (arXiv:1712.04462)

Cai X., Pereyra M., McEwen J. D., 2017a, preprint (arXiv:1711.04818)

Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223

Carrillo R. E., McEwen J. D., Wiaux Y., 2014, MNRAS, 439, 3591

Combettes P. L., Pesquet J. C., 2010, preprint (arXiv:0912.3522v4)

Combettes P. L., Wajs V. R., 2005, Multiscale Model. Simul., 4, 1168

Cornwell T. J., 2008, IEEE Journal of Selected Topics in Signal Processing, 2, 793

Cornwell T. J., Evans K. F., 1985, A&A, 143, 77

Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, A&A, 576, A7

Dabbech A., Wolz L., Pratley L., McEwen J. D., Wiaux Y., 2017, preprint (arXiv:1702.05009)

Durmus A., Moulines E., Pereyra M., 2016, preprint (arXiv:1612.07471v1)

Garsden H., et al., 2015, A&A, 575, A90

Green P. J., Łatuszyński K., Pereyra M., Robert C. P., 2015, Statistics and Computing, 25, 835

Greiner M., Vacca V., Junklewitz H., Enßlin T. A., 2017, preprint (arXiv:1605.04317v2)

Gull S. F., Daniell G. J., 1978, Nature, 272, 686

Högbom J. A., 1974, A&AS, 15, 417

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

Kartik S. V., Carrillo R. E., Thiran J.-P., Y. W., 2017, Springer, New-York, 468, 2382

Li F., Cornwell T. J., de Hoog F., 2011a, A&A, 528, A31

Li F., Brown S., Cornwell T. J., de Hoog F., 2011b, A&A, 531, A126

McEwen J. D., Wiaux Y., 2011, MNRAS, 413, 1318

Moreau J. J., 1965, Bulletin de la Société Mathématique de France, 93, 273

Onose A., Carrillo R. E., Repetti A., McEwen J. D., Thiran J. P., Pesquet J. C., Wiaux Y., 2016, MNRAS, 462, 4314

Onose A., Dabbech A., Wiaux Y., 2017, MNRAS, 469, 938

Pereyra M., 2016a, preprint, (arXiv:1612.06149)

Pereyra M., 2016b, Statistics and Computing, 26, 745

Pereyra M., 2017, SIAM Journal Imaging Sciences, 10, 285

Pereyra M., Bioucas-Dias J., Figueiredo M., 2015, Signal Processing Conference (EUSIPCO)

Pratley L., McEwen J. D., d'Avezac M., Carrillo R. E., Onose A., Wiaux Y., 2018, MNRAS, 473, 1038

Rau U., Bhatnagar S., Voronkov M. A., Cornwell T. J., 2009, Proc. IEEE, 97, 1472

Robert C. P., 2007, The Bayesian Choice. Springer-Verlag New York

Ryle M., Hewish A., 1960, MNRAS, 120, 220

Ryle M., Vonberg D. D., 1946, Nature, 158, 339

Stewart I. M., Fenech D. M., Muxlow T. W. B., 2011, A&A, 535, A81

Suksmono A. B., 2009, Electrical Engineering and Informatics, 1, 110

Sutter P. M., et al., 2014, MNRAS, 438, 768

Thompson A., Moran J., Swenson G., 2017, Interferometry and Synthesis in Radio Astronomy. Springer International Publishing

Wenger S., Magnor M., Pihlström Y., Bhatnagar S., Rau U., 2010, Electrical Engineering and Informatics, 122, 1367

Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009a, MNRAS, 395, 1733

Wiaux Y., Puy G., Boursier Y., Vandergheynst P., 2009b, MNRAS, 400, 1029

Wolz L., McEwen J. D., Abdalla F. B., Carrillo R. E., Wiaux Y., 2013, MNRAS, 436, 1993