
Uncertainty quantification for fast reconstruction methods using augmented equivariant bootstrap: Application to radio interferometry

Mostafa Cherif¹ Tobías I. Liaudat¹ Jonathan Kern¹ Christophe Kervazo² Jérôme Bobin¹

¹IRFU, CEA, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

²LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France

{mostafa.cherif,tobias.liaudat,jonathan.kern,jerome.bobin}@cea.fr
christophe.kervazo@telecom-paris.fr

Abstract

The advent of next-generation radio interferometers like the Square Kilometer Array promises to revolutionise our radio astronomy observational capabilities. The unprecedented volume of data these devices generate requires fast and accurate image reconstruction algorithms to solve the ill-posed radio interferometric imaging problem. Most state-of-the-art reconstruction methods lack trustworthy and scalable uncertainty quantification, which is critical for the rigorous scientific interpretation of radio observations. We propose an unsupervised technique based on a conformalized version of a radio-augmented equivariant bootstrapping method, which allows us to quantify uncertainties for fast reconstruction methods. Noticeably, we rely on reconstructions from ultra-fast unrolled algorithms. The proposed method brings more reliable uncertainty estimations to our problem than existing alternatives.

1 Introduction

Radio interferometry [1] combines signals from multiple antennas to obtain images with very high resolutions. The next-generation radio-interferometry devices, such as the Square Kilometer Array (SKA) [2], will have the potential to unlock scientific discoveries thanks to their unprecedented angular resolution and sensitivity. However, this comes with the computational challenge of processing the incoming data deluge [3]. Interferometric data comprises an incomplete coverage of the Fourier domain (uv-space), which, added to the observational noise, makes the reconstruction problem an ill-posed inverse problem, known as radio interferometric (RI) imaging.

Most families of RI imaging methods, like CLEAN-based [4–6], sparsity-based [7–10], Bayesian approaches [11–14], or learned iterative algorithms [15], suffer from drawbacks such as bad reconstruction quality, due to constraints of the model considered, or long computing times, due to the iterative nature of their algorithm. Learned end-to-end algorithms use recent deep-learning techniques to offer ultra-fast and accurate image reconstructions [16–19]. However, these models lack interpretability, making the attempt to quantify the reconstruction’s uncertainty an arduous task.

It is essential to rigorously quantify our reconstruction’s uncertainties to conduct scientific studies and make decisions based on these reconstructions. The high dimensionality of the RI data makes many standard uncertainty quantification (UQ) techniques, like MCMC sampling-based techniques [20] or conformal prediction [21], impractical or ineffective. Existing UQ methods rely on the Bayesian framework exploiting approximations [22] to avoid sampling [23, 24], but still rely on iterative optimisation techniques and do not exhibit tight bounds on the estimated uncertainty intervals.

Alternative UQ methods rely on Gaussian priors [11, 12] and exploit variational inference [13]. Recent methods, based on ensemble techniques, only quantify the model uncertainty [25, 26], failing to take into account the predominant source of uncertainties that stem from the forward operator’s large null space due to the partial Fourier coverage. Recent methods based on score-based priors [27–29] obtain good results and provide UQ but are not adapted to the large-scale problems raised by SKA-type interferometers as they rely on sampling schemes.

We propose to use a fast and performant learned end-to-end reconstruction method, based on an unrolled architecture, equipped with a conformalized equivariant bootstrapping method adapted to the RI imaging problem. The equivariant bootstrap method is well suited to the RI imaging problem as, by selecting group actions adapted to the forward model’s null space, we can reduce the bootstrapping estimation error, providing us with tight uncertainty bounds and excellent coverage plots. In addition, the subsequent conformalization procedure provides statistical guarantees to the estimated intervals.

2 Radio-interferometric imaging problem

We consider the simplified convolutional form of the observational model for the RI imaging problem, which writes

$$y = M * x^* + n, \quad (1)$$

where $y \in \mathbb{R}^{d \times d}$ represents our observations, i.e., the backprojected visibilities, $M \in \mathbb{R}^{d \times d}$ the point spread function (PSF) models the acquisition process including the incomplete uv-coverage (see Appendix B for examples), $x^* \in \mathbb{R}^{d \times d}$ is the ground truth image, $*$ denotes convolution, and $n \in \mathbb{R}^{d \times d}$ models the observational and instrumental noise, which we suppose, for simplicity, to be white Gaussian additive noise with zero mean and a known standard deviation.

The most computationally expensive operation in RI imaging is applying the forward model due to the (de)gridding operations for the large number of visibilities observed. Therefore, drastically limiting the number of applications of the forward model is mandatory when considering large-scale applications like SKA. Most current RI imaging methods are iterative algorithms, e.g., sparsity-based and Plug-and-Play (PnP) [30], that require many applications of the forward operator for convergence. In this work, we rely on an unrolled architecture [31, 32], where a fixed number of iterations of convex optimization algorithms [33] are unfolded by representing all its operations as layers of a neural network. This choice is computationally efficient allowing it to achieve ultra-fast reconstructions while leveraging the knowledge of the forward operator and obtaining state-of-the-art reconstruction qualities [19]. We use the EVIL-Deconv [19] reconstruction method, where the authors build an unrolled architecture based on LISTA-CP [34] using a DRUNET [35] for denoising. The EVIL-Deconv method uses the PSF as input and the model is trained on a set of PSFs covering various synthetic observational configurations. Therefore, it does not need to be retrained for a change in PSF. A more detailed description of the method can be found in Appendix A.1. Despite the good performance of the method, the reconstruction lacks UQ.

3 Uncertainty Quantification Methods

We start by summarizing the standard parametric bootstrapping procedure. In our forward model Eq. (1), y can be seen as a realization of the probabilistic model, $Y \sim P(M * x^*) := P(\bar{M}x^*)$, where P represents the noise distribution, and \bar{M} a matrix representing the convolution by M . Starting from the reconstructed image $\hat{x}(y)$, we draw bootstrap measurements y_i using the model in Eq. (1) and replacing x^* with $\hat{x}(y)$. Using the same reconstruction method, we then compute $x_i = \hat{x}(y_i)$, which we can compare to $\hat{x}(y)$. From the collection of N bootstrap samples, $\{x_i\}_{i=1}^N$, we build confidence regions for x^* . In this work, we consider the confidence region \mathcal{C}_α using q_α the top α -quantile of the samples $\{|x_i - \hat{x}(y)|\}_{i=1}^N$, with $\mathcal{C}_\alpha = \{x : |x - \hat{x}(y)| < q_\alpha\}$. This conventional parametric bootstrap tends to underestimate the uncertainties in imaging problems where considerable uncertainties span from the forward operator’s large null space [36, §3], as is the case in RI imaging.

Equivariant bootstrap [36] is based on recent developments in the equivariant imaging framework [37, 38] exploiting symmetries in the set of signals \mathcal{X} , with $x^* \in \mathcal{X}$. Let us define \mathcal{G} as a finite group acting on \mathcal{X} with group actions represented by $T_{g_i} \in \mathcal{G}$, an invertible mapping. Assuming that \mathcal{X} is \mathcal{G} -invariant, we can have access to multiple virtual forward operators, $\bar{M}T_{g_i}$, with possibly different null spaces if \bar{M} is not \mathcal{G} -equivariant. The bootstrap method consist of drawing a random group

action $T_{g_i} \in \mathcal{G}$, to then generate a bootstrap measurement \tilde{y}_i from $\tilde{Y}_i \sim P(\bar{M}T_{g_i}\hat{x}(y))$. Later, we compute $\tilde{x}_i = T_{g_i}^{-1}\hat{x}(\tilde{y}_i)$. Finally, we can use the generated bootstrap samples $\{\tilde{x}_i\}_{i=1}^N$ to construct a confidence region \mathcal{C}_α as before, but using \tilde{q}_α the top α -quantile of the samples $\{|\tilde{x}_i - \hat{x}(y)|\}_{i=1}^N$. If T_{g_i} is properly chosen based on the particularities of \mathcal{X} and \bar{M} , the composition $\bar{M}T_{g_i}$ can have different null spaces than \bar{M} helping to probe the variability of the estimator $\hat{x}(Y)$ and characterize its uncertainties with respect to x^* .

3.1 Conformalized Augmented Radio Bootstrap (CARB)

We created a set of group actions calibrated to the RI imaging problem considered for our method, coined Conformalized Augmented Radio Bootstrap (CARB). First, we calibrate and include standard transformations as circular shift transformations not exceeding 2 pixels, image flips over the horizontal and vertical axis, and rotations of 90-degrees multiples, bringing the best results. Based on a convolution, the forward operator \bar{M} is equivariant to translations. However, as the reconstruction operator is not, the composition $\bar{M}T_{g_i}(\hat{x}(y))$ spans a different subspace than $\bar{M}\hat{x}(y)$, making the translation actions useful. We then consider the group of invertible 2D filters in the specific form of low-shelving and high-shelving filters, as shown in Figure 1. These filters attenuate low or high frequencies without cutting them off completely, assuring their inverses exist. The filters prove helpful and adapted to our model as they change the frequency response of the original filter represented by the PSF convolution. The resulting filter, a combination of the shelving filter and the PSF, will probably span a different subspace than the PSF alone. The change in the resulting filter helps to improve the characterisation of the errors from our estimation method, $\hat{x}(Y)$. The probability of applying each of the two filter transformations is 0.5. The high and low drop-off frequencies are selected from two Gaussian distributions around frequencies we have calibrated for the RI imaging problem, with standard deviations of 5. We add a constraint to ensure that the low drop-off frequency stays below the high drop-off frequency. The final group action used to generate the bootstrap samples in CARB is a random composition of the aforementioned transformations, where each transformation is applied with a given probability. This composition allows us to significantly expand the number of possible group actions, helping to estimate uncertainties better.

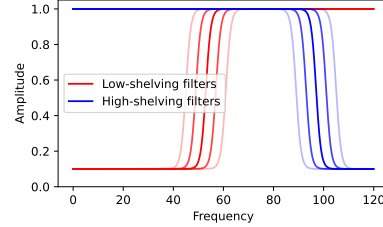


Figure 1: 1D radial visualization of shelving filters.

By constructing the confidence region with the α -quantile, we could expect that for most images, only $100(1 - \alpha)\%$ of their pixels would have absolute residuals above the estimated uncertainty value. Once the confidence regions are estimated, we want to calibrate them to statistically guarantee that the uncertainty intervals around each pixel contain the true value with a particular user-chosen error rate, δ , as is done in [39] on heuristic-based uncertainties. Therefore, we include a subsequent conformalization procedure [21] based on Risk-Controlling Prediction Sets (RCPS) [39], which we describe with more details in Appendix C. In practice, we fix the error rate δ to 0.1 in our numerical experiments.

4 Numerical experiments

We compare the proposed UQ method, CARB, with quantile regression [40], conformalized quantile regression (CQR) [41], a standard parametric bootstrap, the equivariant bootstrap [36] (including rotations, flips, and up to 2-pixel translations), and a version of CARB without conformalization. All the bootstrap-based methods use 500 samples, which was selected based on a study shown in Appendix D. To train our quantile regression method, we keep the same architecture of our unrolling network as the reconstruction, and replace the last layer with two learned neural networks. These networks are trained with the 5th and 95th quantile losses, which is a standard procedure [21, 41].

Data We use 64×64 patches from Hubble space telescope observations as ground truth images to keep the expensive numerical experiments in our computational budget. The simulated PSFs are based on the MeerKAT radio telescope [42] antenna array. We train our unrolled architecture for an SNR range between 30 to 60dB. We use an SNR of 40dB for the rest of the paper and suppose it is known. For all methods comparisons, 10000 images were used to have statistically significant results.

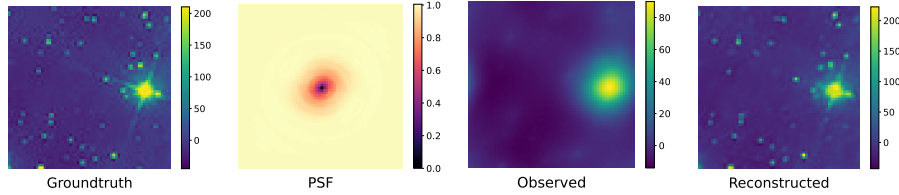


Figure 2: RI image reconstruction result for the unrolling algorithm. From left to right, the figures correspond to: the ground truth observation; the simulated point spread function (PSF); the observed image after the application of the PSF and the observational noise; and the reconstructed image from EVIL-Deconv.

4.1 Results

We show an example of a RI image reconstruction result with the unrolled architecture in Figure 2, with additional results in Appendix E. We then include a quantitative performance comparison with other state-of-the-art reconstruction methods in Appendix A.2.

In Figure 3, we compare the oracle, or ground truth, estimation error with each method’s 90-th quantile estimation to provide a pixel-wise UQ visualization of the reconstruction. We observe a high mismatch between the quantile regression and the oracle error. The conformalization procedure in the CQR tries to correct the mismatch by greatly inflating the quantiles, making them impractical for any interpretation. The equivariant bootstrap-based methods provide error maps with high correlation to the oracle error and tight error bars, which can be further calibrated with the conformalization procedure.

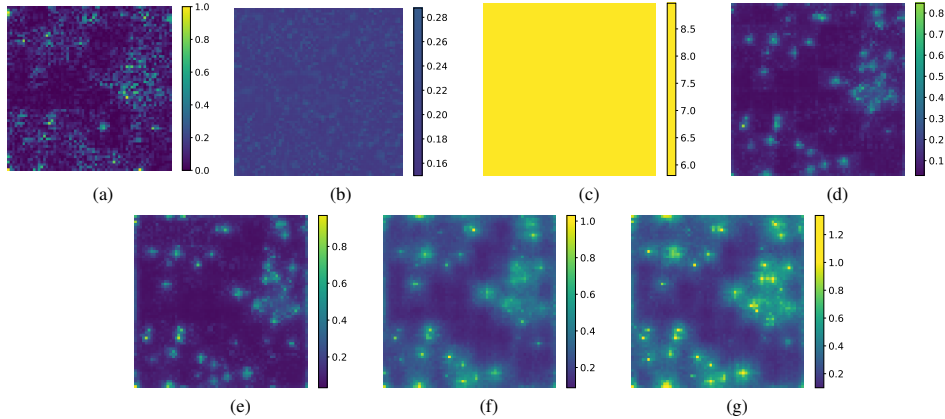


Figure 3: Pixel-wise uncertainty quantification represented by the 90-th quantile estimation. Subfigure (a) shows the oracle absolute residuals. Then, the estimation of each method is presented: (b) quantile regression, (c) conformalized quantile regression, (d) parametric bootstrap, (e) equivariant bootstrap, (f) augmented radio bootstrap and (g) conformalized augmented radio bootstrap.

Table 1: Uncertainty quantification performance comparison based on the average ℓ_2 norm ratio between the 90% confidence interval length and the ground truth image, and the empirical coverage percentage of the interval.

| Method | Length ratio | Coverage |
|---------------------------|--------------|----------|
| Quantile Regression (QR) | 0.15 | 14% |
| Conformalized QR | 204.08 | 92% |
| Parametric Bootstrap | 0.07 | 0% |
| Equivariant Bootstrap | 0.13 | 7% |
| Augmented Radio Bootstrap | 0.29 | 87% |
| CARB | 0.34 | 91% |

The average time to construct confidence regions for each image is around 735ms for all bootstrapping techniques, compared to 153ms for the quantile regressions. The quantile regressions provide reduced times as it is a direct inference of the quantile. The draw of each bootstrap sample is independent of each other once the reconstruction $\hat{x}(y)$ is obtained. This fact allows us to exploit extremely easy parallelization techniques for the bootstrap sample generation, which we currently do by drawing 128 samples in parallel.

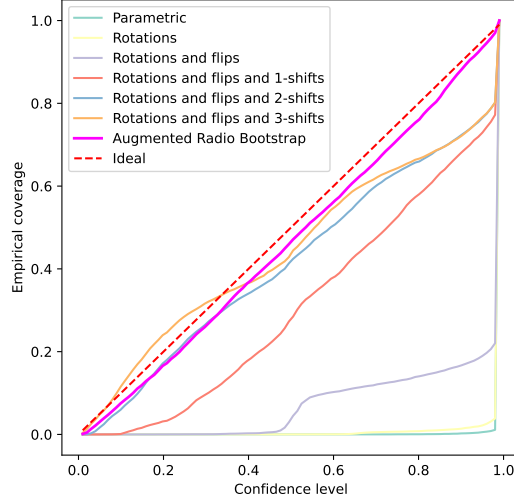


Figure 4: Coverage plots for the equivariant bootstrap methods with different group actions.

A quantitative comparison of the UQ methods is presented in Table 1. We first compare the average ℓ_2 norm ratio between the confidence interval lengths and the ground truth image to study how tight the error estimations are. We then compare an empirical coverage of the confidence interval to verify if they are statistically valid. In other words, we empirically compute the expectation $\mathbb{E}_y\{\mathbf{P}(\hat{I}_\alpha^-(y) < x^* - \hat{x}(y) < \hat{I}_\alpha^+(y))|y\}$ over all pixels in the image dataset, where α is set to 0.1, and $\hat{I}_\alpha^-(y), \hat{I}_\alpha^+(y)$ are the estimated lower and higher interval limits. The result should be greater or equal to $1 - \delta$, with δ being the error rate. The parametric bootstrap significantly underestimates the errors, which is also true for the standard equivariant bootstrap. The error estimation greatly improves when adding the radio-specific transformations, showcasing the importance of setting problem-specific group actions. The subsequent conformalization procedure calibrates the estimated intervals such that the coverage verifies the required $1 - \delta$ probability, allowing CARB to obtain tight bounds with correct coverage. We confirm that the CQR verifies the coverage probability but at the expense of too large uncertainty intervals, making the final UQ less informative.

Lastly, we present in Figure 4 the coverage plots following [36] for the bootstrap-based methods and show how the curve changes as we include different group actions. We estimate a confidence region for x^* , derived from the pivotal statistic $\|x^* - \hat{x}(Y)\|_2^2$ related to the estimation’s mean squared error. We then compute the empirical coverage probabilities on the test set, as measured by the proportion of test images that lie within the confidence regions for a range of confidence levels. The Augmented Radio Bootstrap method obtains an excellent coverage plot, although the probability is slightly underestimated. This result is remarkable in the context of the recent study [43] showing that Bayesian imaging methods are not reporting reliable probabilities for UQ.

One limitation of the current UQ method can be observed when the reconstruction algorithm used catastrophically fails to recover a feature from the ground truth image. In that situation, even if we draw bootstrap samples, the lost feature will not be recovered, and the uncertainty will be misestimated. Another limitation is the need for a representative calibration dataset for the conformalization procedure, which might not always be available in practice. In such a situation, the non-conformalized version of CARB should be used.

Acknowledgments and Disclosure of Funding

This work was performed using HPC resources from GENCI-IDRIS (Grant 2024-AD011015754). The code for CARB, as well as the trained models and the data used in this article, will be made publicly available in a future and more extensive publication.

Software used: numpy [44], Pytorch [45], Matplotlib [46], Jupyter [47], Astropy [48], nenupy [49].

References

- [1] Anthony Thompson, James Moran, and George Swenson, Jr. *Interferometry and Synthesis in Radio Astronomy*. Astronomy and Astrophysics Library. Springer, 01 1991. ISBN 978-3-319-44429-1. doi: 10.1007/978-3-319-44431-4.
- [2] Peter E. F. Dewdney, Peter J. Hall, Richard T. n Schilizzi, and T.J.L.W. Lazio. The square kilometre array. *Proceedings of the IEEE*, 97:1482–1496, 2006.
- [3] A. M. M. Scaife. Big telescope, big data: towards exascale with the square kilometre array. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 378(2166): 20190060, 2020. doi: 10.1098/rsta.2019.0060. URL <https://royalsocietypublishing.org/doi/abs/10.1098/rsta.2019.0060>.
- [4] J. A. Högbom and Tim J. Cornwell. Aperture Synthesis with a Non-Regular Distribution of Interferometer Baselines. *Astronomy and Astrophysics*, 15:417–426, 1974.
- [5] A. R. Offringa, B. McKinley, N. Hurley-Walker, F. H. Briggs, R. B. Wayth, D. L. Kaplan, M. E. Bell, L. Feng, A. R. Neben, J. D. Hughes, J. Rhee, T. Murphy, N. D. R. Bhat, G. Bernardi, J. D. Bowman, R. J. Cappallo, B. E. Corey, A. A. Deshpande, D. Emrich, A. Ewall-Wice, B. M. Gaensler, R. Goeke, L. J. Greenhill, B. J. Hazelton, L. Hindson, M. Johnston-Hollitt, D. C. Jacobs, J. C. Kasper, E. Kratzenberg, E. Lenc, C. J. Lonsdale, M. J. Lynch, S. R. McWhirter, D. A. Mitchell, M. F. Morales, E. Morgan, N. Kudryavtseva, D. Oberoi, S. M. Ord, B. Pindor, P. Procopio, T. Prabu, J. Riding, D. A. Roshi, N. Udaya Shankar, K. S. Srivani, R. Subrahmanyam, S. J. Tingay, M. Waterson, R. L. Webster, A. R. Whitney, A. Williams, and C. L. Williams. WSCLEAN: an implementation of a fast, generic wide-field imager for radio astronomy. *Monthly Notices of the Royal Astronomical Society*, 444(1):606–619, October 2014. doi: 10.1093/mnras/stu1368.
- [6] A. R. Offringa and O. Smirnov. An optimized algorithm for multiscale wideband deconvolution of radio astronomical images. *Monthly Notices of the Royal Astronomical Society*, 471(1):301–316, October 2017. doi: 10.1093/mnras/stx1547.
- [7] J. D. McEwen and Y. Wiaux. Compressed sensing for wide-field radio interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 413(2):1318–1332, 04 2011. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2011.18217.x. URL <https://doi.org/10.1111/j.1365-2966.2011.18217.x>.
- [8] R. E. Carrillo, J. D. McEwen, and Y. Wiaux. Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 426(2):1223–1234, 10 2012. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2012.21605.x. URL <https://doi.org/10.1111/j.1365-2966.2012.21605.x>.
- [9] R. E. Carrillo, J. D. McEwen, and Y. Wiaux. PURIFY: a new approach to radio-interferometric imaging. *Monthly Notices of the Royal Astronomical Society*, 439(4):3591–3604, 02 2014. ISSN 0035-8711. doi: 10.1093/mnras/stu202. URL <https://doi.org/10.1093/mnras/stu202>.
- [10] A Dabbech, A Onose, A Abdulaziz, R A Perley, O M Smirnov, and Y Wiaux. Cygnus A super-resolved via convex optimization from VLA data. *Monthly Notices of the Royal Astronomical Society*, 476(3): 2853–2866, 02 2018. ISSN 0035-8711. doi: 10.1093/mnras/sty372. URL <https://doi.org/10.1093/mnras/sty372>.
- [11] Junklewitz, H., Bell, M. R., Selig, M., and Enßlin, T. A. Resolve: A new algorithm for aperture synthesis imaging of extended emission in radio astronomy. *A&A*, 586:A76, 2016. doi: 10.1051/0004-6361/201323094. URL <https://doi.org/10.1051/0004-6361/201323094>.
- [12] Philipp Arras, Jakob Knollmüller, Henrik Junklewitz, and Torsten A. Enßlin. Radio Imaging With Information Field Theory. *arXiv e-prints*, art. arXiv:1803.02174, March 2018. doi: 10.48550/arXiv.1803.02174.
- [13] Jakob Knollmüller and Torsten A. Enßlin. Metric Gaussian Variational Inference. *arXiv e-prints*, art. arXiv:1901.11033, January 2019. doi: 10.48550/arXiv.1901.11033.
- [14] Philipp Arras, Hertzog L. Bester, Perley, Richard A., Leike, Reimar, Smirnov, Oleg, Westermann, Rüdiger, and Enßlin, Torsten A. Comparison of classical and bayesian imaging in radio interferometry - cygnus a with clean and resolve. *A&A*, 646:A84, 2021. doi: 10.1051/0004-6361/202039258. URL <https://doi.org/10.1051/0004-6361/202039258>.

- [15] Matthieu Terris, Arwa Dabbech, Chao Tang, and Yves Wiaux. Image reconstruction algorithms in radio interferometry: From handcrafted to learned regularization denoisers. *Monthly Notices of the Royal Astronomical Society*, 518(1):604–622, 09 2022. ISSN 0035-8711. doi: 10.1093/mnras/stac2672. URL <https://doi.org/10.1093/mnras/stac2672>.
- [16] Amir Aghabiglou, Chung San Chu, Arwa Dabbech, and Yves Wiaux. The R2D2 Deep Neural Network Series Paradigm for Fast Precision Imaging in Radio Astronomy. *The Astrophysical Journal Supplement Series*, 273(1):3, 2024. doi: 10.3847/1538-4365/ad46f5.
- [17] Matthijs Mars, Marta M. Betcke, and Jason D. McEwen. Learned radio interferometric imaging for varying visibility coverage. *arXiv e-prints*, art. arXiv:2405.08958, May 2024. doi: 10.48550/arXiv.2405.08958.
- [18] Tianlong Chen, Xiaohan Chen, Wuyang Chen, Howard Heaton, Jialin Liu, Zhangyang Wang, and Wotao Yin. Learning to optimize: A primer and a benchmark. *Journal of Machine Learning Research*, 23(189): 1–59, 2022. URL <http://jmlr.org/papers/v23/21-0308.html>.
- [19] Jonathan Kern, Jérôme Bobin, and Christophe Kervazo. EVIL-Deconv: Efficient Variability-Informed Learned Deconvolution using Algorithm Unrolling. [*Manuscript submitted for publication*], 2024.
- [20] Xiaohao Cai, Marcelo Pereyra, and Jason D. McEwen. Uncertainty quantification for radio interferometric imaging - I. Proximal MCMC methods. *MNRAS*, 480(3):4154–4169, November 2018. doi: 10.1093/mnras/sty2004.
- [21] Anastasios Nikolas Angelopoulos and Stephen Bates. A Gentle Introduction to Conformal Prediction and Distribution-Free Uncertainty Quantification. *arXiv e-prints*, art. arXiv:2107.07511, July 2021. doi: 10.48550/arXiv.2107.07511.
- [22] Marcelo Pereyra. Maximum-a-posteriori estimation with bayesian confidence regions. *SIAM Journal on Imaging Sciences*, 10(1):285–302, 2017. doi: 10.1137/16M1071249. URL <https://doi.org/10.1137/16M1071249>.
- [23] Xiaohao Cai, Marcelo Pereyra, and Jason D. McEwen. Uncertainty quantification for radio interferometric imaging: II. MAP estimation. *MNRAS*, 480(3):4170–4182, November 2018. doi: 10.1093/mnras/sty2015.
- [24] Tobías I Liaudat, Matthijs Mars, Matthew A Price, Marcelo Pereyra, Marta M Betcke, and Jason D McEwen. Scalable Bayesian uncertainty quantification with data-driven priors for radio interferometric imaging. *RAS Techniques and Instruments*, 3(1):505–534, 08 2024. ISSN 2752-8200. doi: 10.1093/rasti/rzae030. URL <https://doi.org/10.1093/rasti/rzae030>.
- [25] Matthieu Terris, Chao Tang, Adrian Jackson, and Yves Wiaux. The AIRI plug-and-play algorithm for image reconstruction in radio-interferometry: variations and robustness. *arXiv e-prints*, art. arXiv:2312.07137, December 2023. doi: 10.48550/arXiv.2312.07137.
- [26] Amir Aghabiglou, Chung San Chu, Arwa Dabbech, and Yves Wiaux. R2D2 image reconstruction with model uncertainty quantification in radio astronomy. *arXiv e-prints*, art. arXiv:2403.18052, March 2024. doi: 10.48550/arXiv.2403.18052.
- [27] Noe Dia, M. J. Yantovski-Barth, Alexandre Adam, Micah Bowles, Pablo Lemos, Anna M. M. Scaife, Yashar Hezaveh, and Laurence Perreault-Levasseur. Bayesian Imaging for Radio Interferometry with Score-Based Priors. *arXiv e-prints*, art. arXiv:2311.18012, November 2023. doi: 10.48550/arXiv.2311.18012.
- [28] Yu Sun, Zihui Wu, Yifan Chen, Berthy T. Feng, and Katherine L. Bouman. Provable probabilistic imaging using score-based generative priors. *IEEE Transactions on Computational Imaging*, 10:1290–1305, 2024. doi: 10.1109/TCI.2024.3449114.
- [29] Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. Principled Probabilistic Imaging using Diffusion Models as Plug-and-Play Priors. *arXiv e-prints*, art. arXiv:2405.18782, May 2024. doi: 10.48550/arXiv.2405.18782.
- [30] Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048.
- [31] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML’10*, page 399–406, Madison, WI, USA, 2010. Omnipress. ISBN 9781605589077.

- [32] Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. doi: 10.1109/MSP.2020.3016905.
- [33] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542. URL <https://doi.org/10.1137/080716542>.
- [34] Xiaohan Chen, Jialin Liu, Zhangyang Wang, and Wotao Yin. Theoretical Linear Convergence of Unfolded ISTA and its Practical Weights and Thresholds. *Advances in Neural Information Processing Systems*, 31: arXiv:1808.10038, August 2018. doi: 10.48550/arXiv.1808.10038.
- [35] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-Play Image Restoration with Deep Denoiser Prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, August 2020. doi: 10.48550/arXiv.2008.13751.
- [36] Marcelo Pereyra and Julián Tachella. Equivariant bootstrapping for uncertainty quantification in imaging inverse problems. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pages 4141–4149. PMLR, 02–04 May 2024. URL <https://proceedings.mlr.press/v238/pereyra24a.html>.
- [37] Dongdong Chen, Julián Tachella, and Mike E. Davies. Equivariant imaging: Learning beyond the range space. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4359–4368, 2021. doi: 10.1109/ICCV48922.2021.00434.
- [38] Julián Tachella, Dongdong Chen, and Mike Davies. Sensing theorems for unsupervised learning in linear inverse problems. *J. Mach. Learn. Res.*, 24(1), mar 2024. ISSN 1532-4435. doi: 10.48550/arXiv.2203.12513.
- [39] Anastasios N Angelopoulos, Amit Pal Kohli, Stephen Bates, Michael Jordan, Jitendra Malik, Thayer Alshaabi, Srigokul Upadhyayula, and Yaniv Romano. Image-to-image regression with distribution-free uncertainty quantification and applications in imaging. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 717–730. PMLR, 17–23 Jul 2022. URL <https://proceedings.mlr.press/v162/angelopoulos22a.html>.
- [40] Roger Koenker and Gilbert Bassett. Regression quantiles. *Econometrica*, 46(1):33–50, 1978. ISSN 00129682, 14680262. URL <http://www.jstor.org/stable/1913643>.
- [41] Yaniv Romano, Evan Patterson, and Emmanuel Candes. Conformalized quantile regression. *Advances in neural information processing systems*, 32, 2019. doi: 10.48550/arXiv.1905.03222.
- [42] South African Radio Astronomy Observatory. MeerKAT, 2016. URL <https://public.ska.ac.za/meerkat>.
- [43] David Y. W. Thong, Charlesquin Kemajou Mbakam, and Marcelo Pereyra. Do Bayesian imaging methods report trustworthy probabilities? *arXiv e-prints*, art. arXiv:2405.08179, May 2024. doi: 10.48550/arXiv.2405.08179.
- [44] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. *Nature*, 585(7825): 357–362, September 2020. doi: 10.1038/s41586-020-2649-2. URL <https://doi.org/10.1038/s41586-020-2649-2>.
- [45] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Köpf, Edward Yang, Zach DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. PyTorch: An Imperative Style, High-Performance Deep Learning Library. *arXiv e-prints*, art. arXiv:1912.01703, December 2019. doi: 10.48550/arXiv.1912.01703.
- [46] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

- [47] Thomas Kluyver, Benjamin Ragan-Kelley, Fernando Pérez, Brian Granger, Matthias Bussonnier, Jonathan Frederic, Kyle Kelley, Jessica Hamrick, Jason Grout, Sylvain Corlay, Paul Ivanov, Damián Avila, Safia Abdalla, Carol Willing, and Jupyter development team. Jupyter notebooks ? a publishing format for reproducible computational workflows. In Fernando Loizides and Birgit Schmidt, editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87–90. IOS Press, 2016. URL <https://eprints.soton.ac.uk/403913/>.
- [48] Astropy Collaboration, Adrian M. Price-Whelan, Pey Lian Lim, Nicholas Earl, Nathaniel Starkman, Larry Bradley, David L. Shupe, Aarya A. Patil, Lia Corrales, C. E. Brasseur, Maximilian Nöthe, Axel Donath, Erik Tollerud, Brett M. Morris, Adam Ginsburg, Eero Vaher, Benjamin A. Weaver, James Tocknell, William Jamieson, Marten H. van Kerkwijk, Thomas P. Robitaille, Bruce Merry, Matteo Bachetti, H. Moritz Günther, Thomas L. Aldcroft, Jaime A. Alvarado-Montes, Anne M. Archibald, Attila Bódi, Shreyas Bapat, Geert Barentsen, Juanjo Bazán, Manish Biswas, Médéric Boquien, D. J. Burke, Daria Cara, Mihai Cara, Kyle E. Conroy, Simon Conseil, Matthew W. Craig, Robert M. Cross, Kelle L. Cruz, Francesco D’Eugenio, Nadia Dencheva, Hadrien A. R. Devillepoix, Jörg P. Dietrich, Arthur Davis Eigenbrot, Thomas Erben, Leonardo Ferreira, Daniel Foreman-Mackey, Ryan Fox, Nabil Freij, Suyog Garg, Robel Geda, Lauren Glattry, Yash Gondhalekar, Karl D. Gordon, David Grant, Perry Greenfield, Austen M. Groener, Steve Guest, Sebastian Gurovich, Rasmus Handberg, Akeem Hart, Zac Hatfield-Dodds, Derek Homeier, Griffin Hosseinzadeh, Tim Jenness, Craig K. Jones, Prajwel Joseph, J. Bryce Kalmbach, Emir Karamahmetoglu, Mikołaj Kaluszyński, Michael S. P. Kelley, Nicholas Kern, Wolfgang E. Kerzendorf, Eric W. Koch, Shankar Kulamani, Antony Lee, Chun Ly, Zhiyuan Ma, Conor MacBride, Jakob M. Maljaars, Demitri Muna, N. A. Murphy, Henrik Norman, Richard O’Steen, Kyle A. Oman, Camilla Pacifici, Sergio Pascual, J. Pascual-Granado, Rohit R. Patil, Gabriel I. Perren, Timothy E. Pickering, Tanuj Rastogi, Benjamin R. Roulston, Daniel F. Ryan, Eli S. Rykoff, Jose Sabater, Parikshit Sakurikar, Jesús Salgado, Aniket Sanghi, Nicholas Saunders, Volodymyr Savchenko, Ludwig Schwardt, Michael Seifert-Eckert, Albert Y. Shih, Anany Shrey Jain, Gyanendra Shukla, Jonathan Sick, Chris Simpson, Sudheesh Singanamalla, Leo P. Singer, Jaladh Singhal, Manodeep Sinha, Brigitta M. Sipőcz, Lee R. Spitler, David Stansby, Ole Streicher, Jani Šumak, John D. Swinbank, Dan S. Taranu, Nikita Tewary, Grant R. Tremblay, Miguel de Val-Borro, Samuel J. Van Kooten, Zlatan Vasović, Shresth Verma, José Vinicius de Miranda Cardoso, Peter K. G. Williams, Tom J. Wilson, Benjamin Winkel, W. M. Wood-Vasey, Rui Xue, Peter Yoachim, Chen Zhang, Andrea Zonca, and Astropy Project Contributors. The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *ApJ*, 935(2):167, August 2022. doi: 10.3847/1538-4357/ac7c74.
- [49] Alan Loh, Julien Girard, and the NenuFAR team. nenupy: a python package for the low-frequency radio telescope nenufar, November 2020. URL <https://doi.org/10.5281/zenodo.3667815>.
- [50] Robert Tibshirani. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Methodological)*, 58(1):267–288, 1996. ISSN 00359246. URL <http://www.jstor.org/stable/2346178>.
- [51] Pierre Ablin, Thomas Moreau, Mathurin Massias, and Alexandre Gramfort. Learning step sizes for unfolded sparse coding. *Neural Information Processing Systems*, 32, 2019. doi: 10.48550/arXiv.1905.11071.
- [52] A.N. Tichonov, F. John, and V.J. Arsenin. *Solutions of Ill-posed Problems*. Halsted Press book. V.H. Winston & Sons, 1977.
- [53] Ge Ma, Ziwei Yan, Zhifu Li, and Zhijia Zhao. Efficient iterative regularization method for total variation-based image restoration. *Electronics*, 11(2), 2022. ISSN 2079-9292. doi: 10.3390/electronics11020258.
- [54] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning Deep CNN Denoiser Prior for Image Restoration. *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3929–3938, 2017. doi: 10.48550/arXiv.1704.03264.
- [55] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends® in Optimization*, 1(3): 127–239, 2014. ISSN 2167-3888. doi: 10.1561/24000000003. URL <http://dx.doi.org/10.1561/24000000003>.
- [56] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *Trans. Img. Proc.*, 26(7):3142–3155, jul 2017. ISSN 1057-7149. doi: 10.1109/TIP.2017.2662206. URL <https://doi.org/10.1109/TIP.2017.2662206>.
- [57] Jean-Christophe Pesquet, Audrey Repetti, Matthieu Terris, and Yves Wiaux. Learning Maximally Monotone Operators for Image Recovery. *SIAM J. Imaging Sci.*, 14:1206–1237, 2020. doi: 10.48550/arXiv.2012.13247.

- [58] D.A. Nix and A.S. Weigend. Estimating the mean and variance of the target probability distribution. In *Proceedings of 1994 IEEE International Conference on Neural Networks (ICNN'94)*, volume 1, pages 55–60 vol.1, 1994. doi: 10.1109/ICNN.1994.374138.
- [59] Matthias Rottmann, Pascal Colling, Thomas-Paul Hack, Robin Chan, Fabian Hüger, Peter Schlicht, and Hanno Gottschalk. Prediction error meta classification in semantic segmentation: Detection via aggregated dispersion measures of softmax probabilities, 2019. URL <https://arxiv.org/abs/1811.00648>.

A Radio-interferometry image reconstruction using an unrolled algorithm

A.1 Unrolling method

Finding the image $x^* \in \mathcal{X}$ that minimizes the reconstruction error with the observed radio-interferometry data is an ill-posed problem due to the form of the convolution operator M , or PSF, and the observational noise. Consequently, it is expected to add a regularization term to the optimization problem, which leads to solving a problem of the following form,

$$\hat{x} = \arg \min_{x \in \mathcal{X}} \frac{1}{2} \|y - M * x\|_2^2 + \lambda \mathcal{R}(x), \quad (2)$$

where the regularization \mathcal{R} , can take the form of a simple sparsity constraint using a l_1 norm [50], the total variation function [51–53] or a more complex function by relying on neural-network-based learned denoisers [54, 15].

Several iterative methods have been designed to solve this optimization problem, such as proximal gradient descent algorithms [33], where each iteration can be computed as

$$x^{l+1} = \underbrace{\text{prox}_{\lambda\alpha}}_{\text{regularization}} \left(\underbrace{x^l - \alpha M^*(y - Mx^l)}_{\text{data fidelity gradient}} \right), \quad (3)$$

where $\text{prox}(\cdot)$ is the proximal operator [55] associated with the chosen regularization function and α is the step size. In this work we follow the EVIL-Deconv method [19], and we unroll 10 iterations of Equation (3) resulting in a neural network architecture with 10 layers. Each layer can be expressed as

$$x^{l+1} = p_l(x^l + \Phi_l(M)(y - M * x^l)), \quad (4)$$

where p_l and Φ_l are trainable operators. In practice, p_l is a pretrained DRUNET [35] and Φ_l is a convolutional neural network.

A.2 Unrolling results

In this section, we present performance results obtained using the unrolling architecture described in the previous subsection, EVIL-Deconv [19], and compare it with two iterative methods, the CLEAN algorithm [4] and a Plug-and-Play (PnP) method, where the denoiser, a DnCNN [56], is trained with a non-expansiveness constraint following [15] to ensure convergence [57]. We study two reconstruction performance metrics: the Normalized Mean Squared Error (NMSE), which we define as follows

$$\text{NMSE [-dB]} = -20 \log_{10} \left(\frac{\|x^* - \hat{x}\|_2}{\|x^*\|_2} \right), \quad (5)$$

and the Structural Similarity Index Measure (SSIM). The mean reconstruction time for each method is also included to show the computational advantage of algorithm unrolling over iterative methods. Table 2 presents the performance results, which were computed on the data described in Section 4. All numerical experiments were done on two Nvidia Tesla V100 SXM2 16Gb GPUs.

Table 2: Performance comparison of RI image reconstruction algorithms.

| Method | Median NMSE [-dB] \uparrow | SSIM \uparrow | Time [ms] \downarrow |
|-------------|------------------------------|-----------------|------------------------|
| CLEAN | 4.2 | 0.296 | 794 |
| PnP | 16.3 | 0.869 | 1110 |
| EVIL-Deconv | 19.9 | 0.970 | 51 |

B Point spread function and uv-coverage

The uv-coverages used are based on the MeerKAT [42] observatory antenna positions. To simulate the PSFs, we use code from the nenupy [49] package. We present examples of uv-coverages and their associated PSFs in Figure 5.

C Conformalization

We suppose we have access to a reconstruction function $\hat{x}(\cdot)$ that maps an observed image $y \in \mathbb{R}^{d \times d}$ and a PSF $M \in \mathbb{R}^{d \times d}$ to a reconstructed image $\hat{x}(y, M)$. Our task is to create uncertainty intervals around each

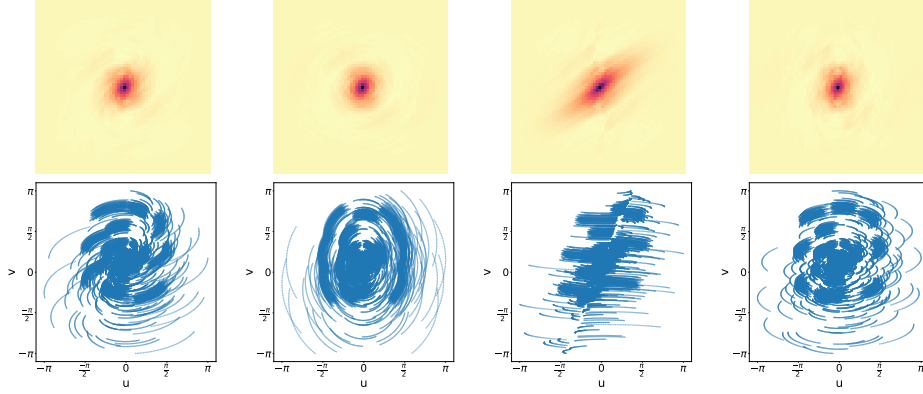


Figure 5: Examples of the simulated MeerKAT PSFs and their associated uv-coverage.

reconstructed image pixel. These intervals should contain the true pixel values with a user-specified probability. Formally, we construct, for each pixel (m, n) , the interval

$$\mathcal{I}_{(m,n)}(y, M) = \left[\hat{x}(y, M)_{(m,n)} - \tilde{l}_{(m,n)}(y, M), \hat{x}(y, M)_{(m,n)} + \tilde{u}_{(m,n)}(y, M) \right], \quad (6)$$

where \tilde{l} and \tilde{u} are heuristic lower and upper interval lengths. These heuristics can, for example, be pixel-wise Gaussian standard deviations [58], softmax output distributions [59], pixel-wise quantiles [40], or residual magnitude regressions. In this work, we assume $\tilde{l} = \tilde{u}$, and that they are both equal to the pixel-wise 90-th quantile of the sampled bootstrap absolute residuals (see Section 3). The uncalibrated intervals are heuristic because they do not contain the ground truth with the desired probability, as we made no assumptions about the algorithm used to train \tilde{l} and \tilde{u} .

In Risk-Controlling Prediction Set (RCPS) [39], the objective is to calibrate the interval $\mathcal{I}_{(m,n)}(y, M)$ such that

$$P(\mathbb{E}[\mathcal{S}(X, \mathcal{I}_{(m,n)}(y, M))] > \alpha) \leq \delta, \quad (7)$$

where \mathcal{S} is an interval score defined as

$$\mathcal{S}(X_i, \mathcal{I}_{(m,n)}(y_i, M_i)) = 1 - \frac{|\{(m, n) : X_{i(m,n)} \in \mathcal{I}_{(m,n)}(y_i, M_i)\}|}{d * d}, \quad (8)$$

which is the proportion of pixels where the ground truth is contained in the constructed intervals. The inner expectation in Equation 7 is over a new test point, while the outer probability is over the calibration set. Once calibrated, the intervals $\mathcal{I}_{(m,n)}(y, M)$ will be an (α, δ) -RCPS. Let us suppose that we dispose of a calibration set $(X_i, Y_i, M_i)_{i \in [1; N_C]}$, where N_C is the calibration set size and X refers to the ground truth image we are hoping to reconstruct. Our goal is to calibrate the intervals using conformal prediction such that they become RCPS verifying Equation 7. We can obtain conformalized intervals as follows

$$\mathcal{I}_{(m,n)}^{(\lambda)}(y, M) = \left[\hat{x}(y, M)_{(m,n)} - \lambda * \tilde{l}_{(m,n)}(y, M), \hat{x}(y, M)_{(m,n)} + \lambda * \tilde{u}_{(m,n)}(y, M) \right], \quad (9)$$

where λ should be chosen such that if $\alpha \in [0, 1]$ is a user-selected risk level and $\delta \in [0, 1]$ is a user-selected error level, then, the new intervals contain at least $1 - \alpha$ of the ground truth pixel values with probability $1 - \delta$. If we set $\alpha = \delta = 0.1$, then at least 90% of the ground truth images should have at least 90% of their pixels within the constructed intervals.

In practice, we take $N_C = 1000$ images with their corresponding PSFs. We then determine the heuristic intervals $\mathcal{I}_i^{(\lambda)} = \mathcal{I}_{(m,n)}^{(\lambda)}(y_i, M_i)$ for each element of this set. Later, we compute a score for each of these intervals $\mathcal{S}_i(\lambda) = \mathcal{S}(X_i, \mathcal{I}_i^{(\lambda)})$ using Equation 8. In order to pick the smallest $\hat{\lambda}$ satisfying our goal, we follow [39] and form Hoeffding's upper-confidence bound

$$H(\lambda) = \frac{1}{N_C} \sum_{i=1}^{N_C} \mathcal{S}_i(\lambda) + \sqrt{\frac{1}{2N_C} \log \left(\frac{1}{\delta} \right)}. \quad (10)$$

We can then pick $\hat{\lambda} = \{\min_{\lambda} : H(\lambda) > \alpha\}$.

D Choice of the number of bootstrap samples

We must choose the number of samples to use in the bootstrapping techniques. To accomplish this, we study the average variation of the norms of pixel-wise uncertainty images as we increase the number of samples. In

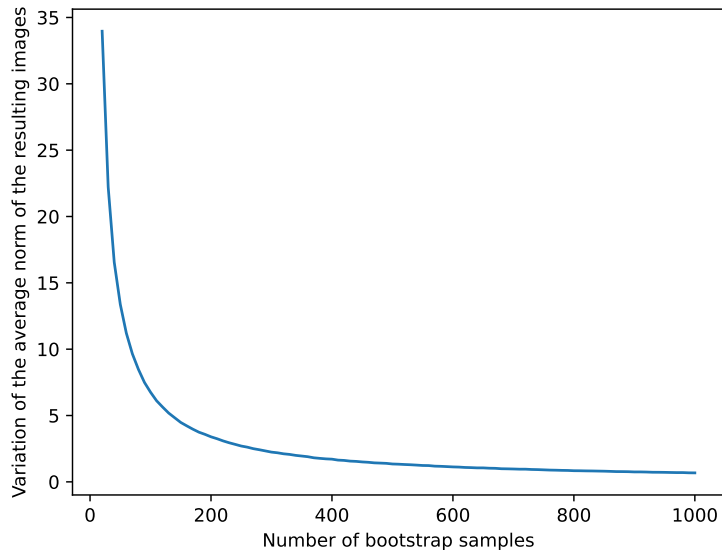


Figure 6: Variation of the 90-th quantile estimation in CARB when increasing the number of bootstrap samples.

practice, we use the 90-th quantile error estimations described in Section 3. Figure 6 presents the curve obtained. We select a number of samples such that the curve is close to stagnating, which means that adding more samples to the bootstrap will not significantly change our confidence regions. This choice allows us to ensure sufficient samples while minimizing the run time of our algorithm.

E Additional results

Figure 7 shows additional astronomical images with their reconstructions and two methods to visualize the estimated CARB uncertainties. The first one takes the 90-th quantile of each pixel of our sampled estimated errors, as shown in Figure 3 for all bootstrap methods. The second alternative way of visualizing the pixel-wise uncertainties is to take the standard deviation of the computed bootstrap samples, as is done in [36].

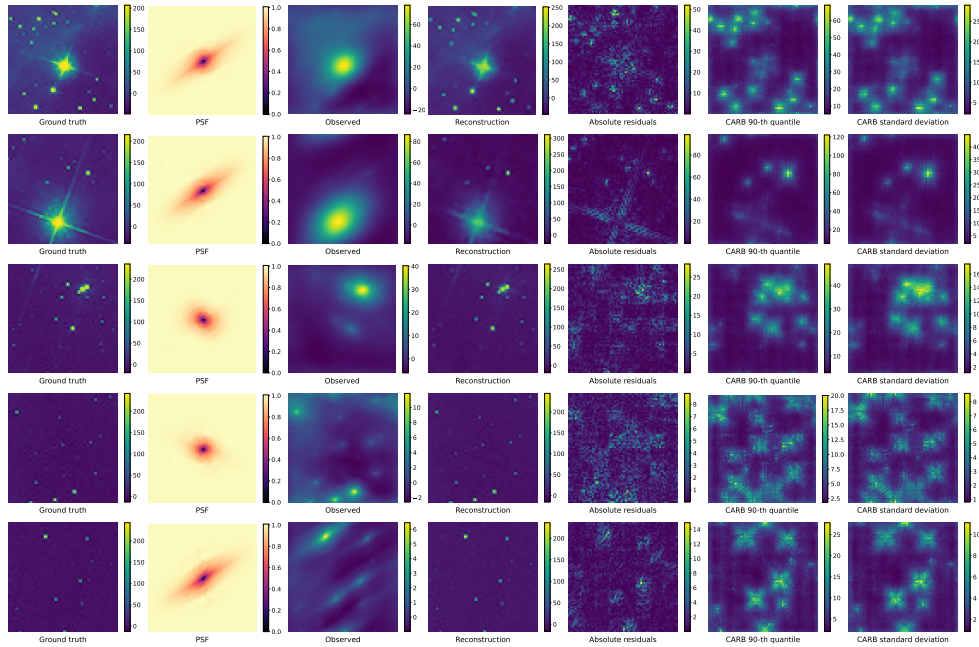


Figure 7: Additional radio interferometric image unrolling algorithm reconstruction results with uncertainty quantification from CARB. The first two columns show the ground truth images used with their associated PSF. The third column presents the observation, and the fourth column is the EVIL-Deconv reconstruction. The fifth column presents the oracle, or ground truth, absolute residuals. The last two columns introduce the pixel-wise uncertainty estimated from the CARB using the 90-th quantile and the standard deviation of the bootstrap samples.