

# Direction-dependent calibration with image-domain gridding

Sebastian van der Tol<sup>1,\*</sup>, Sarod Yatawatta<sup>1</sup>, Bram Veenboer<sup>1</sup>, and David Rafferty<sup>2</sup>

<sup>1</sup> Netherlands Institute for Radio Astronomy (ASTRON), Postbus 2, 7990 AA Dwingeloo, The Netherlands

<sup>2</sup> University of Hamburg, Gojenbergsweg 112, 21029 Hamburg, Germany

Received 26 November 2025 / Accepted 27 January 2026

## ABSTRACT

**Aims.** Wide-field images made by radio interferometers are invariably affected by direction-dependent systematic effects such as the ionosphere or the beam pattern. Calibration along a set of discrete directions in the sky is the default technique to estimate and correct these systematic errors. However, additional processing such as smoothing and mosaicing are required to reconcile the step wise variation of the estimated systematic errors at the edges of the discrete directions (facets).

**Methods.** We overcome the discrete nature of direction-dependent calibration by using image-domain gridding as the model for the calibration. Instead of discrete directions in the sky, calibration is performed using a basis that represents a set of sub-grids in the Fourier space. This automatically removes the need for extra operations to recover the wide-field systematics error model without any discontinuity.

**Results.** We provide results based on LOFAR data where we compare the traditional facet-based (discrete directional gains) calibration with the proposed approach. The comparison shows improved image quality, mainly because of the physical plausibility of the proposed approach as opposed to using a piecewise constant model for direction-dependent systematic errors.

**Key words.** methods: numerical – techniques: interferometric

## 1. Introduction

Radio interferometers that have wide fields of view, for example the low frequency array (LOFAR; van Haarlem et al. 2013) and the Murchison wide-field array (MWA; Tingay et al. 2013), require specialized data processing for the estimation and correction of direction-dependent systematic errors across their observing fields. Direction-dependent calibration is an essential step in this data-processing chain, and there are a plethora of methods that accomplish this such as Kazemi et al. (2011), Yatawatta (2015), van Weeren et al. (2016), and Tasse et al. (2018). The application of corrections for these systematic errors during image synthesis is an equally important step in the data processing chain, and there are also many wide-field imaging and deconvolution methods that accomplish this, including Bhatnagar et al. (2013), Bhatnagar et al. (2008), Tasse et al. (2013), Offringa et al. (2014), and van der Tol et al. (2018).

While there are advanced methods for (i) direction-dependent calibration and for (ii) wide-field imaging with correction for direction-dependent errors, they are mostly disjointed. In other words, methods and software that fall into either of these categories have been developed with a focus exclusively on direction-dependent calibration or on wide-field imaging with correction for direction-dependent errors. A typical image synthesis from raw data is performed by first running direction-dependent calibration (estimating systematic errors) and thereafter running image synthesis (correcting for systematic errors in the data). Direction-dependent calibration is performed along discrete directions in the sky, with the estimated systematic errors only being applicable to a small area in the sky centered around this direction. This area is normally called a facet. After making images of each facet using image-synthesis

techniques that incorporate direction-dependent effects, such facets are combined using techniques such as mosaicing (Sault et al. 1996) to build the full field-of-view image. With minor deviations, the aforementioned scheme is the core data-processing strategy currently used by LOFAR imaging (van Weeren et al. 2016; Tasse et al. 2018).

The unification of direction-dependent calibration and imaging has seen limited growth compared to the disjointed methods that specialize in either calibration or imaging. This is mainly due to the complexity of joint calibration and imaging in a computationally feasible manner. Most joint methods in calibration with imaging parameterize both the sky and the instrument systematics and find a joint solution. Due to the restriction in computational complexity, existing methods that fall into the category of joint direction-dependent calibration and imaging are few in number, can only handle simple physical models for the systematics, or can only handle narrow fields of view (e.g., Intema et al. 2009; Arras et al. 2019; Birdi et al. 2020; Roth et al. 2023). In this work, we overcome most of these limitations by proposing the use of image-domain gridding (IDG; van der Tol et al. 2018) as the forward model prediction in direction-dependent calibration. In comparison to existing methods for joint calibration and imaging, the proposed method is different because we still have separate calibration (instrument-parameterized) and imaging (sky-parameterized) steps. During the calibration step, we estimated the direction-dependent errors covering the full field of view, parameterized using an arbitrary basis in space. The estimated A-terms (or systematic error terms) are thereafter directly applied during the image synthesis, without having an intermediate systematic-error model creation step. In order to reduce the computational complexity in calibration, we use stochastic optimization methods (Yatawatta et al. 2019) that can minimize optimization problems with a large parameter space with super-linear convergence. A noteworthy similarity

\* Corresponding author: [tol@astron.nl](mailto:tol@astron.nl)

of the proposed method partially exists with the method by Roth et al. (2023) in the use of IDG for forward model prediction. However, the proposed method is computationally more efficient than that of Roth et al. (2023) because we follow a frequentist approach (maximum likelihood) rather than a Bayesian approach, and therefore our method is compatible with wide fields of view.

The remainder of this paper is organized as follows. In Section 2 we give an overview of IDG and how it is used in calibration (IDG-CAL). Next, in Section 3 we describe the stochastic optimization scheme used in calibration. In Section 4 we provide results based on LOFAR observations to illustrate the efficacy of IDG-CAL compared to the conventional facet-based calibration. We discuss the computational and physical advantages of the proposed method compared to alternatives in Section 5. Finally, we draw our conclusions in Section 6.

Please note that lowercase bold letters refer to column vectors (e.g.,  $\mathbf{y}$ ). Uppercase bold letters refer to matrices (e.g.,  $\mathbf{C}$ ). Indices are lower case letters. Where possible, the number of elements over which an index runs is indicated by the same letter in uppercase. Unless otherwise stated, all parameters are complex numbers. The set of complex numbers is given as  $\mathbb{C}$ , and the set of real numbers is given as  $\mathbb{R}$ . The imaginary unit is denoted by  $j$  (a dotless  $j$ ). The matrix Hermitian transpose is referred to as  $(\cdot)^H$ . The matrix Kronecker product is given by  $\otimes$ , and the matrix Hadamard product is given by  $\odot$ . The identity matrix of size  $N$  is given by  $\mathbf{I}_N$ . The Frobenius norm is given by  $\|\cdot\|$ . The function  $\text{vec}(\cdot)$  stacks the columns of a matrix in a vector, and  $\text{diag}(\cdot)$  makes a diagonal matrix from a vector.

## 2. Data model

In this section, we provide a brief introduction to IDG (van der Tol et al. 2018; Veenboer et al. 2017) and how it is used as the forward model prediction for direction-dependent calibration.

### 2.1. Image-domain gridding

Image domain gridding (van der Tol et al. 2018) computes the equivalent of A-projection (Bhatnagar et al. 2008) while avoiding the computation of over-sampled convolution kernels. This is advantageous in situations where many different kernels are needed for relatively few visibilities per kernel.

With regard to correction for direction-dependent errors in image synthesis, both A-projection and IDG offer the advantage of applying a continuous correction over facet-based methods that apply piecewise constant corrections. However, since traditional direction-dependent calibration methods are facet-based, and interpolation from facet-based corrections to an A-term introduces new errors, the A-projection and related methods have (so far) mostly been used to apply analytical beam models only, for which the A-term can be readily computed.

The calibration method IDG-CAL presented in this paper is based on the realization that IDG is sufficiently efficient in applying varying A-terms and that it is feasible to use IDG iteratively in a calibration scheme to compute derivatives and residuals. This allows us to solve for A-terms directly from the visibilities, instead of from intermediate solutions.

Per baseline, the visibilities are partitioned in time-frequency ranges,  $T_p$ , such that the uv-coverage convolved by a kernel fits within sub-grid  $p$ . The kernel includes a taper, the A-term, and a differential w-term. A sub-grid consists of  $Q = N_g \times N_g$  pixels (typically  $32 \times 32$ ). For efficiency reasons,  $N_g$  is always chosen

to be a multiple of eight. Each sub-grid is extracted from the Fourier transform of the full model image and then transformed back to the image domain. A sub-grid belongs to a single baseline. Per baseline, there is at least one sub-grid, but there can be more. Sub-grids can overlap and be placed arbitrarily on the uv (or Fourier space) grid.

Once the sub-grids are known, only a single equation remains to compute the model visibilities. The calibration algorithm presented in this paper was derived from the equation provided below. The model visibilities,  $\mathbf{V}_{jk}$  ( $\in \mathbb{C}^{2 \times 2}$ ), for baseline  $j$  and time-frequency index  $k$  are given as a sum over the pixels of sub-grid  $p$  as

$$\mathbf{V}_{jk} = \sum_{q=1}^Q \mathbf{A}_{s_{1p}q} \mathbf{V}_{pq} \mathbf{A}_{s_{2p}q}^H \varphi_{pqk}, \quad (1)$$

where  $s_{1p}$  and  $s_{2p}$  are the station (or antenna) indices (total  $S$  stations) corresponding to baseline  $j_p$  associated with sub-grid  $p$ . In the context of LOFAR data processing, a station is essentially an electronically steerable antenna. The value of pixel  $q$  of sub-grid  $p$  in the image domain is denoted by  $\mathbf{V}_{pq}$ . The A-terms representing the systematic errors of stations  $s_{1p}$  and  $s_{2p}$  on sub-grid  $p$ , evaluated at pixel  $q$ , are given by  $\mathbf{A}_{s_{1p}q}$  and  $\mathbf{A}_{s_{2p}q}$ , respectively. Note that  $\mathbf{A}_{s_{1p}q}, \mathbf{A}_{s_{2p}q}, \mathbf{V}_{pq} \in \mathbb{C}^{2 \times 2}$ . The complex exponential phase term is given by  $\varphi_{pqk}$ . This term represents the propagation delay, which depends on the sub-grid pixel position  $(l_q, m_q)$  and the  $u_{pk}, v_{pk}, w_{pk}$  coordinates of the visibility  $\mathbf{V}_{jk}$ .

Note that form (1) is exactly the same as the faceted approach. The only difference here is that the summation is over sub-grid pixels and not facets. Effectively, IDG applies a correction that is interpolated from sub-grid resolution to full image resolution. This approaches physical reality much closer than what can be obtained using a facet-based approach.

### 2.2. Parameterization of A-terms

An A-term in IDG is sampled at sub-grid resolution. To limit the number of degrees of freedom, the A-terms are written as an expression of a few parameters using an appropriate basis. The parameterization is rather arbitrary, since IDG-CAL does not depend on the precise choice of the basis as long as the derivatives with respect to the parameters can be computed.

A simple example of a parameterization (a scalar A-term) can be given as

$$\mathbf{A}_q(\mathbf{x}) = \left( \sum_{r=1}^{R_a} a_r h_r(l_q, m_q) \right) \exp \left( j \sum_{r=1}^{R_b} b_r h_r(l_q, m_q) \right) \mathbf{I}_2, \quad (2)$$

where  $j$  denotes the imaginary unit ( $j^2 = -1$ ). The polynomial basis is expressed as  $h_1(l, m) = 1$ ,  $h_2(l, m) = l$ ,  $h_3(l, m) = m$ ,  $h_4(l, m) = l^2$ ,  $h_5(l, m) = lm$ ,  $h_6(l, m) = m^2$ , and so on. In Eq. (2),  $q$  is the sub-grid pixel index and  $l_q$  and  $m_q$  are the image coordinates of that pixel. The parameters,  $\mathbf{x}$ , are partitioned into parts for the amplitude,  $\mathbf{a}$ , and for the phase,  $\mathbf{b}$ :

$$\mathbf{x} = \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix}, \quad (3)$$

where  $\mathbf{a}$  is a vector of length,  $R_a$ , and  $\mathbf{b}$  is a vector of length,  $R_b$ , with real parameters. The total number of parameters for all  $P$  sub-grids and  $S$  stations is therefore  $R = (R_a + R_b) \times S$ . Note that a more detailed parameterization can be used to model diagonal A-terms or A-terms for full polarization, but we left out this detail for simplicity.

The input data are the vector,  $\mathbf{y}$ , containing all observed visibilities that are being calibrated. The function  $\hat{\mathbf{V}}_{jk}(\mathbf{y})$  selects the visibilities corresponding to the baseline,  $j$ , and time-frequency index,  $k$ , from input  $\mathbf{y}$  and returns a matrix ( $\in \mathbb{C}^{2 \times 2}$ ). Likewise, the model visibilities can be written as function of all the free parameters as  $\mathbf{V}_{jk}(\mathbf{x})$ . This function only truly depends on a subset of  $\mathbf{x}$ , i.e., the entries corresponding to baseline  $j$  and time frequency index  $k$ . The model visibilities also depend on the model image, through the sub-grids. This dependence is not included in the notation because the model image is not a free parameter that is going to be solved for here in this formulation. The difference between the observed visibilities and the model visibilities are the residual visibilities given by

$$\tilde{\mathbf{V}}_{jk}(\mathbf{x}, \mathbf{y}) = \hat{\mathbf{V}}_{jk}(\mathbf{y}) - \mathbf{V}_{jk}(\mathbf{x}). \quad (4)$$

The goal of the calibration algorithm is to find the parameters that minimize a cost function measuring the error in terms of the norm of the residual. The cost function to be minimized is designed as a weighted sum of squares of the residual visibilities (under a Gaussian noise model, summed over  $J$  baselines and  $K$  time-frequency samples):

$$f(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^J \sum_{k=1}^K \text{vec}(\tilde{\mathbf{V}}_{jk}(\mathbf{x}, \mathbf{y}))^H \bar{\mathbf{W}}_{jk} \text{vec}(\tilde{\mathbf{V}}_{jk}(\mathbf{x}, \mathbf{y})), \quad (5)$$

where  $\bar{\mathbf{W}}_{jk} = \text{diag}(\text{vec}(\mathbf{W}_{jk}))$  contain the data weights (given a priori).

Note that any given sub-grid belongs to a single baseline and a time-frequency selection. Since there are  $P$  sub-grids in total, the cost function can be split into a sum over per-sub-grid cost functions as

$$f(\mathbf{x}, \mathbf{y}) = \sum_{p=1}^P C_p(\mathbf{x}, \mathbf{y}). \quad (6)$$

The derivative of the cost function (5) can be calculated in closed form by considering the per-sub-grid cost function (6) with respect to parameter  $x_{sr}$  for station  $s$  and parameter index  $r$  as

$$\frac{\partial}{\partial x_{sr}} f(\mathbf{x}, \mathbf{y}) = \sum_{p \in P_s} \frac{\partial}{\partial x_{sr}} C_p(\mathbf{x}, \mathbf{y}), \quad (7)$$

where  $P_s$  is a set of indices to sub-grids in which station  $s$  participates.

### 2.3. Per-sub-grid equations of the derivative

In order to calculate the closed form derivative, we considered the separable, sub-grid-based cost function (6). Without loss of generality, we only considered a single sub-grid,  $p$ , for simplicity. The associated time-frequency range is denoted by  $T_p$ , the baseline  $j_p$ , and the stations  $s_{1p}$  and  $s_{2p}$ .

The cost function associated with sub-grid  $p$  is given by

$$C_p(\mathbf{x}, \mathbf{y}) = \sum_{k \in T_p} \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y}))^H \bar{\mathbf{W}}_{j_p k} \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y})). \quad (8)$$

The derivative of the per-sub-grid cost function is given by

$$\frac{\partial}{\partial x_{sr}} C_p(\mathbf{x}, \mathbf{y})$$

$$\begin{aligned} &= \sum_{k \in T_p} \left( \text{vec} \left( \frac{\partial}{\partial x_{sr}} \tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y}) \right)^H \bar{\mathbf{W}}_{j_p k} \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y})) \right. \\ &\quad \left. + \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y}))^H \bar{\mathbf{W}}_{j_p k} \text{vec} \left( \frac{\partial}{\partial x_{sr}} \tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y}) \right) \right) \\ &= 2 \text{Re} \sum_{k \in T_p} \text{vec} \left( \frac{\partial}{\partial x_{sr}} \tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y}) \right)^H \bar{\mathbf{W}}_{j_p k} \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y})) \\ &= -2 \text{Re} \sum_{k \in T_p} \text{vec} \left( \frac{\partial}{\partial x_{sr}} \mathbf{V}_{j_p k}(\mathbf{x}, \mathbf{y}) \right)^H \bar{\mathbf{W}}_{j_p k} \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y})). \quad (9) \end{aligned}$$

The derivative with respect to  $x_{s_{1p}r}$  of the model visibilities is given by

$$\frac{\partial}{\partial x_{s_{1p}r}} \mathbf{V}_{j_p k} = \sum_{q=1}^Q \frac{\partial}{\partial x_{s_{1p}r}} \mathbf{A}_{s_{1p}q} \mathbf{V}_{pq} \mathbf{A}_{s_{2p}q}^H \varphi_{pqk}, \quad (10)$$

for all  $k \in T_p$ . The derivative with respect to the parameters for the other station,  $x_{s_{2p}r}$ , is given in similar fashion, but is not worked out here. In  $\text{vec}(\cdot)^H$  form, this becomes

$$\frac{\partial}{\partial x_{s_{1p}r}} \text{vec}(\mathbf{V}_{jk})^H = \sum_{q=1}^Q \text{vec}(\mathbf{V}_{pq})^H \left( \mathbf{A}_{s_{2p}q}^* \otimes \frac{\partial}{\partial x_{s_{1p}r}} \mathbf{A}_{s_{1p}q} \right)^H \varphi_{pqk}^*. \quad (11)$$

Substituting Eq. (11) into Eq. (9) gives

$$\begin{aligned} \frac{\partial}{\partial x_{s_{1p}r}} C_p(\mathbf{x}, \mathbf{y}) &= -2 \text{Re} \sum_{k \in T_p} \sum_{q=1}^Q \left( \text{vec}(\mathbf{V}_{pq})^H \right. \\ &\quad \left. \times \left( \mathbf{A}_{s_{2p}q}^* \otimes \frac{\partial}{\partial x_{s_{1p}r}} \mathbf{A}_{s_{1p}q} \right)^H \varphi_{pqk}^* \bar{\mathbf{W}}_{j_p k} \text{vec}(\tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y})) \right). \quad (12) \end{aligned}$$

Then, we define the weighted residual as

$$\mathbf{X}_{pq} = \sum_{k \in K_p} \varphi_{j_p k q}^* \mathbf{W}_{j_p k} \odot \tilde{\mathbf{V}}_{j_p k}(\mathbf{x}, \mathbf{y}). \quad (13)$$

Rearranging Eq. (12) and writing it in terms of the accumulated residual visibilities,  $\mathbf{X}_{pq}$ , this becomes

$$\frac{\partial}{\partial x_{s_{1p}r}} C_p(\mathbf{x}, \mathbf{y}) = -2 \text{Re} \sum_{q=1}^Q \text{vec}(\mathbf{V}_{pq})^H \text{vec} \left( \frac{\partial}{\partial x_{s_{1p}r}} \mathbf{A}_{s_{1p}q}^H \mathbf{X}_{pq} \mathbf{A}_{s_{2p}q} \right). \quad (14)$$

The derivation above leads to Algorithm 1 for the computation of the derivatives.

Note that this procedure is analogous to the way the residual image is computed with IDG. This analogy is not surprising since the residual image is the derivative of cost function (5) with respect to the pixel values of the model image (Rau et al. 2009).

When only the A-term has changed, the updated derivative can be computed by starting from step 5. In Section 3, we discuss how the calculated derivative is used to minimize the calibration cost function (5).

**Table 1.** Variation of computational time and final error with various configuration parameters.

Iterations	Epochs	Minibatches	Total time	Time per iteration	Normalized error
48	1	1	4590	95	1.0000
8	3	2	1614	26	1.0156
5	3	3	1028	17.5	1.0251
4	3	4	892	14.5	1.0250
3	3	5	668	12.5	1.0356
3	3	6	663	9.5	1.0446
3	3	7	617	7.5	1.0393

**Algorithm 1** Derivative computation**Require:** Model image, sub-grid size

- 1: Fourier transform the model image to the uv-domain.
- 2: Split off sub-grids from uv-grid.
- 3: Fourier transform sub-grids to the image domain.
- 4: Apply the known effects (beam, smearing) to the sub-grids.
- 5: Compute the model visibilities from the sub-grids according to (1) and subtract them from the observed visibilities according to (4) to obtain the residual visibilities.
- 6: Accumulate the residual visibilities onto sub-grids as per (13).
- 7: Compute the per-sub-grid contribution to the derivatives per (14) and accumulate them to obtain the total derivative.

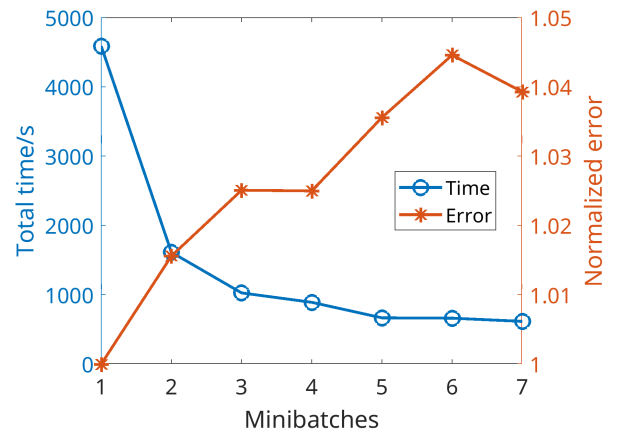
**3. Stochastic calibration**

In this section, we describe the optimization strategy used for IDG-CAL for computational efficiency and stability. We used the closed form expressions for the cost function (5) as well as its derivative as given in Algorithm 1 for this purpose. We considered calibration using IDG to be equivalent to minimizing the nonlinear, non-convex cost function  $f(\mathbf{x}, \mathbf{y})$  (5):

$$\mathbf{x} = \arg \min_{\mathbf{x}} f(\mathbf{x}, \mathbf{y}), \quad (15)$$

where  $\mathbf{x}$  are the parameters and  $\mathbf{y}$  are the data. At a local minimum, the gradient of the cost  $\nabla f(\mathbf{x}, \mathbf{y}) = \mathbf{0}$ . Therefore, in order to find the local minimum, an essential ingredient is the gradient or the derivative. In stochastic optimization, this gradient is approximated (Robbins & Monro 1951) using a subset of the data  $\tilde{\mathbf{y}}$ , i.e.,  $\nabla f(\mathbf{x}, \mathbf{y}) \approx \nabla f(\mathbf{x}, \tilde{\mathbf{y}})$ . The point of this approximation is that the cost of computing  $\nabla f(\mathbf{x}, \tilde{\mathbf{y}})$  is much less than the cost of computing  $\nabla f(\mathbf{x}, \mathbf{y})$ .

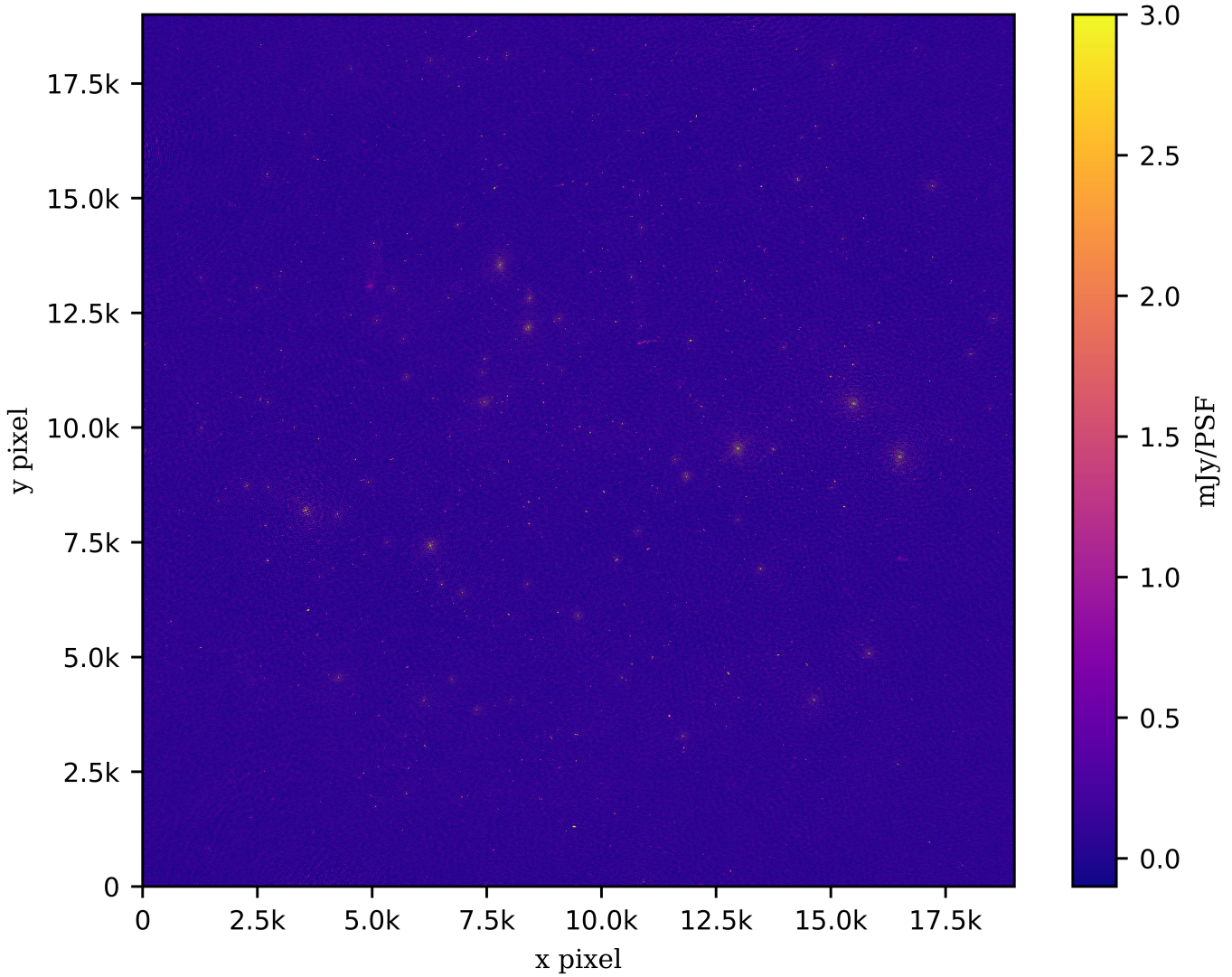
For solving problems similar to (15), we developed a stochastic version of the limited-memory Broyden Fletcher Goldfarb Shanno (LBFGS; Liu & Nocedal 1989; Byrd et al. 1995) algorithm using this idea of stochastic approximation (Yatawatta et al. 2019; Yatawatta 2020). In addition to the reduced computational cost, by using the LBFGS algorithm, we were able to achieve faster (super-linear) convergence compared to popular first-order optimization methods based on gradient descent (Yatawatta 2020). We applied the same algorithm for IDG-CAL, the only exception being that instead of using subsets of data, we used subsets of sub-grids (sub-grids are explained in Section 2) to approximate the gradient (and the cost function). It is also possible to divide data into subsets according to the frequency (when there are multiple frequencies or separate channels), but we leave this optimization for future work. The overall stochastic calibration scheme is shown in Algorithm 2. In this algorithm,

**Fig. 1.** Total time and final error variation with the number of mini-batches. A large number of mini-batches implies a small mini-batch size, which, hence, reduces the computational cost. The error is normalized by the error reached with the full-batch mode of solving.

we used the term mini-batch to denote a subset of data (according to sub-grids or channels). An epoch is defined as a sequence of mini-batches that completes the full set of data being calibrated. With each mini-batch of data, a small number of iterations of the LBFGS algorithm is performed. The total amount of data is divided into the number of given mini-batches, and therefore, in each iteration, the amount of data being used is smaller than the total amount of data being calibrated. However, since one epoch includes the full number of mini-batches, during one epoch, the total amount of data is indeed used to obtain the calibration solutions.

In Table 1, we show the timing and final error results of a typical calibration with varying mini-batch and epoch sizes. This example uses data with 30 channels that are divided into blocks of ten (three blocks), and the sub-grid size is  $48 \times 48$ , making the total numbers of sub-grids equal to 2304 (image size is  $8k \times 8k$  pixels).

In Fig. 1, we illustrate the variation of the total computation time and the final error with the mini-batch size (values extracted from Table 1). We note two distinct behaviors in Fig. 1. First, we should remember that the amount of data (or sub-grids) used by a single minibatch is inversely proportional to the number of mini-batches. In Fig. 1, the  $x$ -axis shows the number of mini-batches increasing from one to seven. When the number of mini-batches is equal to one, we have the original full-batch mode of optimization. In full-batch mode, all data are used in the evaluation of (15) per iteration, resulting in the highest computational cost (or computational time). In contrast, as the number of mini-batches increases, the amount of data (or sub-grids) that are being used



**Fig. 2.** Full field-of-view image produced by standard facet-based calibration after four major cycles. The image size is  $19k \times 19k$  pixels of  $1.25'' \times 1.25''$  covering a field of view of about  $7 \times 7$  square degrees. The color-scale is linear within the range of  $[-0.1, 3]$  mJy/PSF.

---

#### Algorithm 2 Stochastic optimization

---

**Require:** Epochs ( $E$ ), mini-batches ( $M$ ), number of iterations ( $K$ )

```

1: epoch = 0
2: for epoch <  $E$  do
3:   minibatch = 0
4:   for minibatch <  $M$  do
5:     iteration = 0
6:     for iteration <  $K$  do
7:       Update parameters by minimizing (15) using the
       selected minibatch
8:       iteration++
9:     end for
10:    minibatch++
11:  end for
12:  epoch++
13: end for
14: Output solution

```

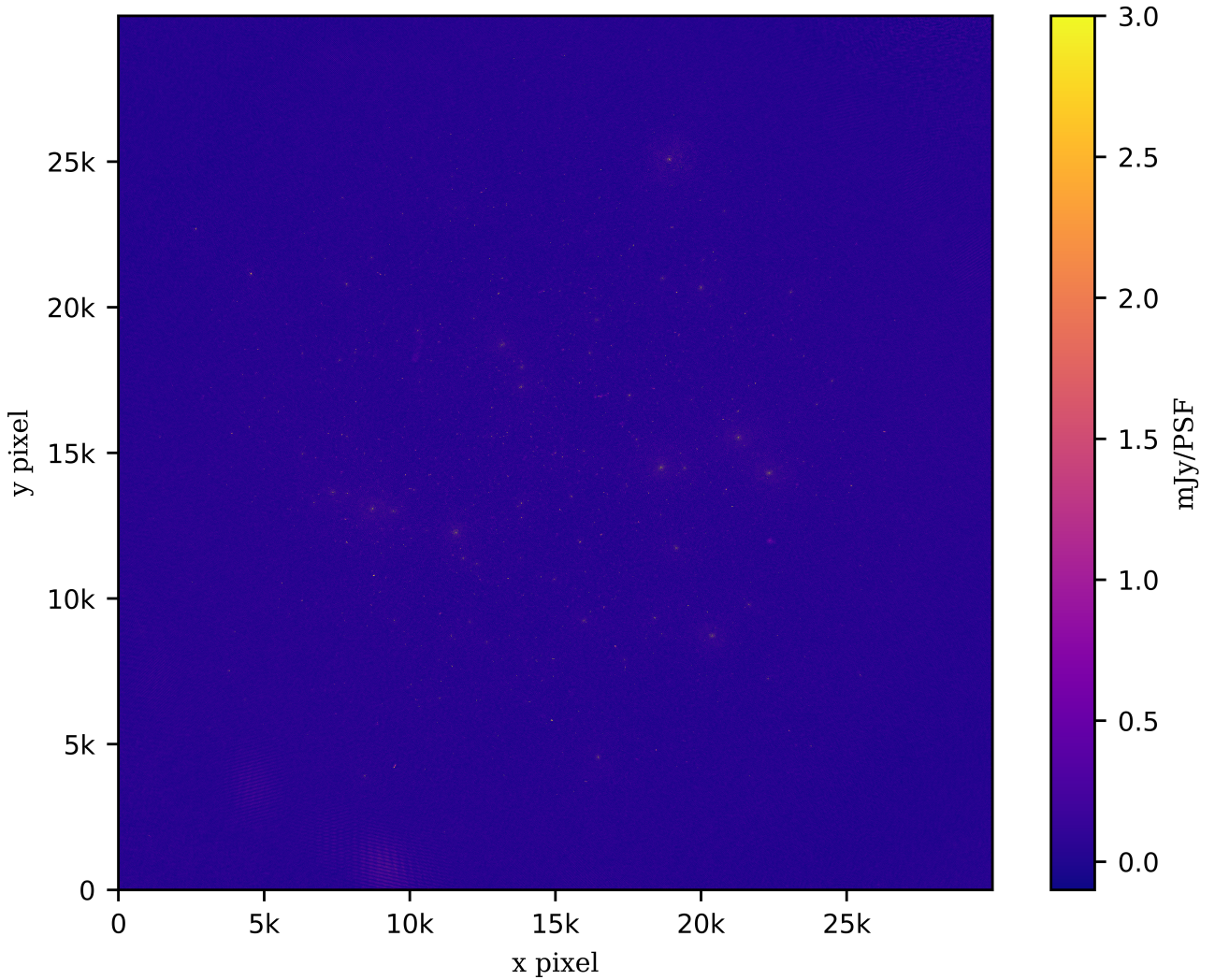
---

to evaluate Eq. (15) and its derivative per iteration is reduced, hence reducing the computational time.

An unfavorable consequence of using mini-batches to solve Eq. (15) is the increase in the final error (or the final residual). However, as seen in Fig. 1, this increase is very low (a few percent) compared to the error attained by the full-batch mode of calibration. We reached a compromise between the computational cost and the error by using a mini-batch size of 3 or 4, which is what we used in our tests discussed in Section 4.

## 4. Results

To illustrate and compare the performance of the proposed calibration scheme, we used a LOFAR high-band array (HBA) observation that is typical of standard LOFAR HBA observations, both in terms of observational setup and of radio frequency interference and ionospheric conditions. The data were observed in 2017 as part of the LC9 019 program. The target (NGC 6338) was observed by LOFAR for eight hours using 48 MHz of bandwidth in the HBA-low band (centered at  $\approx 144$  MHz). The target observation was preceded by a ten-minute calibrator observation of 3C380. The calibrator and target data were obtained from the LOFAR long-term archive (stored with SAS IDs 632471 and 632475 for the calibrator and target data, respectively) and



**Fig. 3.** Full field-of-view image produced by IDG-CAL after three major cycles. The image size is 30k×30k pixels of  $1.2'' \times 1.2''$  covering a field of view of about  $10 \times 10$  square degrees. The color-scale is linear within the range of  $[-0.1, 3]$  mJy/PSF (the same as in Fig. 2).

processed with version 5.0 of the LOFAR initial calibration (LINC) pipeline to remove instrumental effects and perform a basic direction-independent calibration on the target field. The data were then averaged to 8 s per time slot and 0.1 MHz per frequency channel, as is typical for HBA data (e.g., Shimwell et al. 2017). International LOFAR stations were not included in the processing. For the tests described below, a subset of 125 frequency channels ( $\approx 12$  MHz), centered at  $\approx 144$  MHz, was used.

In Fig. 2, we show the image produced by facet-based imaging as implemented in the Raptor pipeline. This image was made after running four major cycles where each one involves a sky-model update, direction-dependent calibration (using DDE-Cal), and facet-based imaging. There are 34 facets covering the field of view of about  $7 \times 7$  square degrees.

For comparison, in Fig. 3, we show the result obtained by IDG-CAL after running only three major cycles (instead of four major cycles) using the same data. Since we were not restricted by the number of facets, we were able to image a larger  $10 \times 10$  square degrees in Fig. 3. For a facet-based approach, imaging a larger field of view would require increasing the number of facets used, which would thus increase the computational cost and make the solutions less constrained (more degrees of freedom). In order to run IDG-CAL, we used  $32 \times 32$  sub-grids with

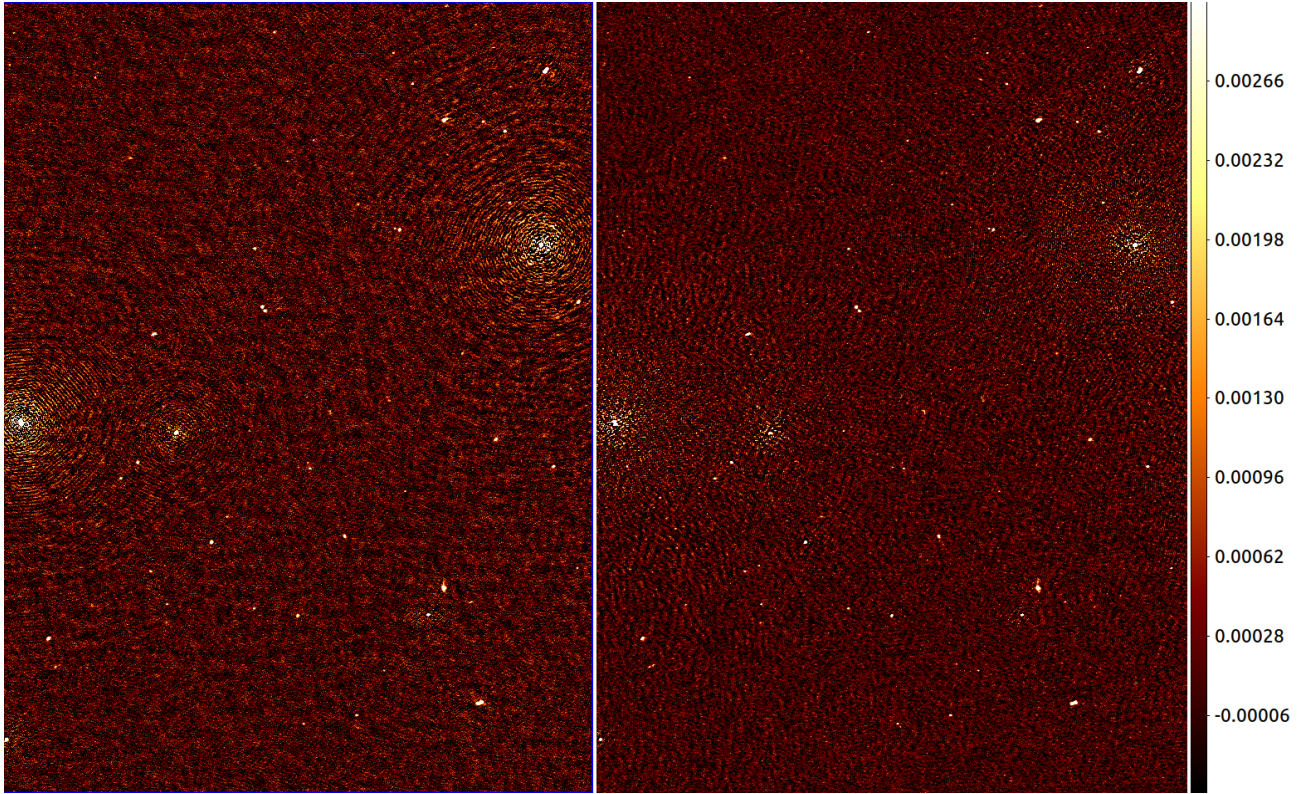
fourth-order polynomials to model the amplitude and phase in Eq. (2). The number of degrees of freedom for both facet-based calibration and for IDG-CAL are almost the same in this comparison. Note that the artifacts at the bottom left edge of Fig. 3 are mainly due to the limitations of the chosen basis functions and the tapering parameters. Such artifacts can be minimized by fine-tuning the aforementioned parameters.

In order to make a more qualitative comparison, we show enlarged parts of both Figs. 2 and 3 covering the same region in Fig. 4. Both panels in Fig. 4 have the same color-scale. The left panel in Fig. 4 was made using facet-based calibration, while the right panel in Fig. 4 is made using IDG-CAL.

We clearly see that the IDG-CAL image in the right panel of Fig. 4 has fewer artifacts around the bright sources. The background noise level is also about 30% lower in the IDG-CAL image for this reason. In Section 5, we discuss the differences between two approaches in more detail.

## 5. Discussion

In this section, we discuss the unique characteristics of IDG-CAL in comparison to traditional facet-based calibration as well as its limitations and potential future improvements. The



**Fig. 4.** Closeup view of a small part ( $1 \times 1.3$  square degrees) of the images from facet calibration using Rapthor (left) and that from IDG-CAL (right). The color-scale is in Jy/PSF and is shown on the right hand side and is the same for both panels.

scalability of calibration algorithms is critical for next-generation LOFAR processing. IDG-CAL offers a more scalable approach than the traditional facet-based DDECAL (e.g., Smirnov & Tasse 2015) by representing direction-dependent effects as a compact image-domain basis rather than as many independent directional solutions.

In DDECAL, calibration is done per facet, with each facet requiring a separate model prediction of visibilities. This causes data duplication, high memory usage, and limited computational intensity. Runtime grows with the number of facets, which increases steeply for wide-field, high-resolution imaging. In the context of the LOFAR enhanced network for sharp surveys (LENSS), which expands LOFAR’s instantaneous field of view by  $4\times$ , this facet approach becomes impractical as the number of facets and the replicated model data grow rapidly.

In contrast, IDG-CAL models the direction dependent effects using image-domain polynomials defined on sub-grids. The computational cost scales roughly linearly with the number of polynomial parameters, not with the number of directions. The method is computation-bound rather than memory-bound; dense kernels dominated by trigonometric and complex arithmetic define time-to-solution methods. These kernels map onto graphics processing units (GPUs) efficiently; their special function units handle trigonometric operations effectively. Modern GPUs deliver a large energy-efficiency advantage for this workload (e.g.,  $\sim 223$  GFLOPs/W for a Blackwell GPU versus  $\sim 5.6$  GFLOPs/W for a Haswell CPU).

As seen in Section 4, IDG-CAL can reach calibration quality comparable to a full facet-based calibration with about three major cycles. The present CPU implementation of IDG-CAL is only slower than DDECAL by about a factor of four, but the algorithm’s parallel structure makes it well-suited to GPU acceleration. GPU speedups of about  $10\times$  over an

optimized CPU implementation have been demonstrated for IDG (Veenboer et al. 2017). We therefore expect similar speed improvements of IDG-CAL on accelerator hardware, making it computationally viable.

For LENSS, which increases processed data volume by  $4\times$  and targets images up to  $100k \times 100k$  pixels, both methods scale with data volume. IDG-CAL preserves predictable computational scaling and modest memory use by increasing sub-grid size and polynomial order. DDECAL would require many more facets, causing large increases in model replication, memory footprint, and I/O.

Summarizing the computational advantages, IDG-CAL scales linearly with model complexity, maps efficiently to GPUs, and avoids the memory and I/O bottlenecks of facet-based calibration. These properties align with LENSS goals for sustainable, high-resolution, wide-field calibration.

The expectation is that a parameterized continuous screen would match physical reality better than the piece-wise constant correction with facets. Evidence that this is indeed the case is given by the background noise in the IDG-CAL self-calibration run compared to a facet-based run in Fig. 4. The Rapthor run uses 24, 31, 33, and 34 facets for the four consecutive self-calibration cycles. In contrast, the IDG-CAL run uses second-, third-, and fourth-order polynomials; i.e., 6, 10, and 15 degrees of freedom in three self-calibration cycles. As seen in Fig. 4, in the third major cycle, with a fourth-order polynomial with 15 degrees of freedom covering an area of  $10 \times 10$  square degrees, we obtain a lower background noise ( $345 \mu\text{Jy}$ ) than the fourth cycle with 34 facets covering an area of  $7 \times 7$  square degrees ( $445 \mu\text{Jy}$ ) with traditional facet-based calibration. Summarizing the quality advantage, we find that the physical plausibility of the A-term-based model for the direction-dependent systematic errors provides a model that consumes

fewer degrees of freedom, as well as a model that preserves the continuity and, for the most part, the differentiability throughout the full field of view.

We present the first results based on IDG-CAL using real data in this work. We have many future improvements planned that we briefly list below:

- GPU acceleration. As discussed above, the IDG-CAL approach is amenable to GPU acceleration. We will explore this together with the use of reduced-precision floating-point calculations when feasible;
- Frequency-dependent sky model. So far, we have used model images that are made for narrow frequency bands. We will explore creating frequency-dependent model images where the pixels have spectral information as well as the intensity. Alternatively, we will explore the use of a stack of model images for different frequencies that can be interpolated for the desired frequency;
- Better basis functions. We will explore the use of more refined basis functions for modeling the A-term;
- Spectral regularization. We will explore the use of spectral regularization (Yatawatta 2015) to further improve the quality of calibration by constraining the frequency variation of the parameters.

The only limitation of IDG-CAL is its intimate link to an image. In facet-based calibration, the sky model can have discrete components and can work with sources that are far apart without creating an image covering all sources that are present. A way to overcome this limitation is to subtract the strong outlier sources from the data (that are not of scientific interest in order to make images) before running IDG-CAL.

## 6. Conclusions

We introduce IDG-CAL, an alternative method for direction-dependent calibration of radio interferometric data using a physically plausible A-term based model. In comparison to traditional facet-based calibration, we see improved image quality as well as the possibility for computational efficiency especially when imaging large fields of view. Further improvements of IDG-CAL are being planned and will improve both the quality of the calibration solutions as well as make the computations run even more efficiently.

*Acknowledgements.* We thank the reviewer and editor for their comments. This work has been supported by FuSE: Fundamental Sciences e-Infrastructure, Dutch Research Council (NWO) Roadmap Project 184.035.004. LOFAR, the Low Frequency Array designed and constructed by ASTRON, has facilities in several countries, owned by various parties (each with their own funding sources), and collectively operated by the International LOFAR Telescope (ILT) foundation under a joint scientific policy. The software used in this paper are: IDG (<https://git.astron.nl/RD/idg>), DP3 (<https://git.astron.nl/RD/DP3>) (van Diepen et al. 2018), RAPHOR (<https://git.astron.nl/RD/raphor>), SAGECAL (<https://sagecal.sourceforge.net/>), and WSClean (Offringa et al. 2014).

## References

- Arras, P., Frank, P., Leike, R., Westermann, R., & Enßlin, T. A. 2019, *A&A*, **627**, A134
- Bhatnagar, S., Cornwell, T. J., Golap, K., & Uson, J. M. 2008, *A&A*, **487**, 419
- Bhatnagar, S., Rau, U., & Golap, K. 2013, *ApJ*, **770**, 91
- Birdi, J., Repetti, A., & Wiaux, Y. 2020, *MNRAS*, **492**, 3509
- Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. 1995, *SIAM J. Sci. Comput.*, **16**, 1190
- Intema, H. T., van der Tol, S., Cotton, W. D., et al. 2009, *A&A*, **501**, 1185
- Kazemi, S., Yatawatta, S., Zaroubi, S., et al. 2011, *MNRAS*, **414**, 1656
- Liu, D. C., & Nocedal, J. 1989, *Math. Program.*, **45**, 503
- Offringa, A. R., McKinley, B., et al. 2014, *MNRAS*, **444**, 606
- Rau, U., Bhatnagar, S., Voronkov, M. A., & Cornwell, T. J. 2009, *IEEE Proc.*, **97**, 1472
- Robbins, H., & Monroe, S. 1951, *Ann. Math. Statist.*, **22**, 400
- Roth, J., Arras, P., Reinecke, M., et al. 2023, *A&A*, **678**, A177
- Sault, R. J., Staveley-Smith, L., & Brouw, W. N. 1996, *A&AS*, **120**, 375
- Shimwell, T. W., Röttgering, H. J. A., Best, P. N., et al. 2017, *A&A*, **598**, A104
- Smirnov, O. M., & Tasse, C. 2015, *MNRAS*, **449**, 2668
- Tasse, C., van der Tol, S., van Zwieten, J., van Diepen, G., & Bhatnagar, S. 2013, *A&A*, **553**, A105
- Tasse, C., Hugo, B., Mirmont, M., et al. 2018, *A&A*, **611**, A87
- Tingay, S. J., Goeke, R., Bowman, J. D., et al. 2013, *PASA*, **30**, e007
- van der Tol, S., Veenboer, B., & Offringa, A. R. 2018, *A&A*, **616**, A27
- van Diepen, G., Dijkema, T. J., & Offringa, A. 2018, DPPP: Default Pre-Processing Pipeline, Astrophysics Source Code Library [[record ascl:1804.003](https://ui.adsabs.org/abs/2018ascl.1804003)]
- van Haarlem, M. P., Wise, M. W., Gunst, A. W., et al. 2013, *A&A*, **556**, A2
- van Weeren, R. J., Williams, W. L., Hardcastle, M. J., et al. 2016, *ApJS*, **223**, 2
- Veenboer, B., Petschow, M., & Romein, J. W. 2017, in *2017 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, 545
- Yatawatta, S. 2015, *MNRAS*, **449**, 4506
- Yatawatta, S. 2020, *MNRAS*, **493**, 6071
- Yatawatta, S., De Clercq, L., Spreeuw, H., & Diblen, F. 2019, in *2019 IEEE Data Science Workshop (DSW)*, 208