CLUSTERED RADIO INTERFEROMETRIC CALIBRATION

S. Kazemi\(^1\), S. Yatawatta\(^1,2\), S. Zaroubi\(^1,3\)

\(^1\) Kapteyn Astronomical Institute, Univ. of Groningen, The Netherlands
\(^2\) ASTRON, The Netherlands
\(^3\) Technion, Physics Department, Haifa 32000, Israel

Abstract
This paper introduces an amendment to radio interferometric calibration of sources below the noise level. The main idea is to employ the information of the stronger sources’ measured signals as a plug-in criterion to solve for the weaker ones. For this purpose, we construct a number of source clusters, with centroids mainly near the strongest sources, assuming that the signals of the sources belonging to a single cluster are corrupted by almost the same errors. Due to this characteristic of clusters, each cluster is calibrated as a single source, using all the coherencies of its sources simultaneously. The obtained solutions for every cluster are assigned to all the cluster’s sources. An illustrative example reveals the superiority of this calibration compared to the un-clustered calibration.

Index Terms— Calibration, Clustering methods, Clustering algorithms, Interferometry: Radio interferometry

1. INTRODUCTION

Calibration of radio synthesis arrays refers to the estimation and reduction of errors introduced by the atmosphere and the incorporated instruments, before imaging. It is the most crucial task in order to achieve the interferometer’s desired precision and sensitivity. Early radio astronomy used external (classical) calibration which is based on estimating the instrument unknown parameters by a celestial radio source with known properties. The external calibration is then improved by self-calibration \(^1\), which utilizes only the observed data for estimating both the source and instrumental unknowns.

Although the calibration of radio telescopes highly benefits from various self-calibration techniques, its performance in interferometric source subtraction is still limited to sources that have a high enough Signal to Noise Ratio (SNR) to be distinguished from the background noise \(^2\) \(^3\) \(^4\). The novelty of this paper is that the presented method has a high performance in source calibration below the noise level, utilizing the strongest sources’ signals. The implementation of such a calibration, termed as “clustered calibration”, is performed by clustering the sources during the calibration process. The clustered calibration improves the information used for calculating solutions by incorporating the total of signals observed at each cluster instead of each individual source’s signal. Thus, in the case of calibrating the low signals of very weak sources, it provides a considerably better result compared with the un-clustered calibration.

Clustering, from the data mining point of view, can be defined as the gathering of similar data points together into groups. An overview of different clustering methods is given by \(^5\). We intend to cluster radio sources in the way that visibilities of sources belonging to a single cluster are affected by almost the same errors, and subsequently could share the same calibration’s solutions. This assumption is valid when the clusters’ angular diameters are small enough so that the variation in their actual solutions is negligible. After arranging the cluster’s centroids mainly near the strongest sources, we calibrate considering every cluster as a single source, and assign the obtained solutions to all the cluster’s sources.

We present the data model of clustered calibration and apply two clustering methods: (i) weighted K-means \(^6\) \(^7\) and (ii) divisive hierarchical clustering \(^8\), to cluster the sky’s sources. In an illustrative example, we demonstrate the superiority of the clustered calibration compared to the un-clustered calibration, in solving for the sources below the noise level, using data observed by the LOw Frequency ARray (LOFAR) synthesis radio telescope \(^9\).

The following notations are used in this paper: Bold, lower case letters refer to column vectors, e.g., \(y\). Upper case bold letters refer to matrices, e.g., \(C\). All parameters are complex numbers, unless stated otherwise. The transpose and conjugation of a matrix are presented by \((\cdot)^T\) and \((\cdot)^*\), respectively. The matrix Kronecker product, the set membership, the empty set, and the union operator are denoted by \(\otimes\), \(\in\), \(\emptyset\), and \(\cup\), respectively.

2. CLUSTERED CALIBRATION DATA MODEL

In this section, we briefly describe the data model of the radio interferometric clustered calibration. For more details on radio interferometry the reader is referred to \(^9\) and for the data

---

The authors acknowledge the Netherlands Foundation for Scientific research (NWO) grant 436040 for support and SZ thanks the Lady Davis grant.

\(^1\) http://www.lofar.org
model of radio interferometric calibration to [10] [11] [12].

Consider an interferometric array consisting of \( N \) receivers, each with two orthogonal polarized feeds \( X \) and \( Y \). The induced voltages at the \( p \)-th receiver’s feeds, \( \tilde{v}_{pi} = [\tilde{v}_{Xpi} \, \tilde{v}_{Ypi}]^T \), due to the polarized waves radiated by the \( i \)-th source, \( e_i = [e_{Xi} \, e_{Yi}]^T \), is given by

\[
\tilde{v}_{pi} = \tilde{J}_{pi} e_i.
\]  
(1)

In Eq. (1), \( \tilde{J}_{pi} \) represents the 2 \( \times \) 2 Jones matrix [10], corresponding to corruptions of \( e_i \) signal at receiver \( p \), which is a product of different Jones matrices as [12]

\[
\tilde{J}_{pi} \equiv G_p E_{pi} Z_{pi} F_{pi} K_{pi}.
\]  
(2)

In Eq. (2), \( K_{pi} \), \( E_{pi} \), \( Z_{pi} \), and \( F_{pi} \) are the Fourier transform, antenna’s voltage pattern, ionospheric phase fluctuation, and Faraday Rotation matrices corresponding to the \( i \)-th source’s direction and receiver \( p \)-th location on the earth, respectively, and \( G_p \) is the receiver \( p \)-th clock phase and electronic gain.

We assume that the total of signals seen at each receiver is a superposition of the \( K \) sources’ corrupted signals, plus the receiver’s thermal noise. Note that the multitude of the ignored fainter sources also contribute to the noise. After correcting the geometric delays corresponding to receivers’ locations on the earth, we correlate the collected signals at every pair of receivers to obtain visibilities [10]. Since the complex gain of Jones matrix \( G \) does not depend on the source direction, it is initially calculated at every receiver and then the visibilities are corrected for it. Stacking up all the corrected visibilities in vector \( y \), we arrive to the general data model of [13] [14] as

\[
y = \sum_{i=1}^{K} s_i + n.
\]  
(3)

In Eq. (3), \( n \) is the additive noise vector, normally assumed to be Gaussian white noise. The nonlinear function \( s_i \) shows the contribution of the \( i \)-th source in the observation:

\[
s_i \equiv \begin{bmatrix} J_{2i} \otimes J_{1i} \text{vec}(C_{(12)i}) \\ \vdots \\ J_{N_i} \otimes J_{(N-1)i} \text{vec}(C_{(N-1)N_i}) \end{bmatrix},
\]

where

\[
J_{pi} \equiv E_{pi} Z_{pi} F_{pi}, \quad C_{(pq)i} \equiv K_{(pq)i} C_i.
\]  
(4)

\( C_i \) is the source’s coherency matrix [15] [10], and the scalar Jones matrix \( K_{(pq)i} \) corresponds to Fourier transform between the source’s direction and the baseline \( pq \).

The calibration is essentially an estimation of the \( J \) Jones matrix, and the removal of the \( K \) brightest sources. However, in practice, the \( E, Z \), and \( F \) Jones matrices obtained for nearby directions and for a given receiver are almost the same. Thus, for every receiver \( p \), if the \( i \)-th and \( j \)-th sources have a small angular separation from each other we have

\[
\tilde{J}_{pi} \approx \tilde{J}_{pj}.
\]  
(5)

Eq. (5) is the underlying assumption for clustered calibration and it tells us that the Fourier transform \( K \) Jones matrix is the only Jones matrix which should be calculated individually for all directions.

Assume that we have \( Q \) (\( Q \ll K \)) source clusters \( L_i \), for \( i \in \{1, \ldots, Q\} \), with small enough angular diameters. Based on Eq. (5), for every cluster \( L_i \), we define

\[
C'_{(pq)i} \equiv \sum_{l \in L_i} C_{(pq)l},
\]  
(6)

and by substituting this new definition in Eq. (3), we formulate the clustered calibration data model, where the index \( i \) is over the clusters and not separate sources. Various techniques can be used to solve this non-linear data model and one of the more popular of them is the Least Squares (LS) optimization algorithm which is discussed along other methods in [13] [14].

### 3. CLUSTERING OF RADIO SOURCES

Suppose that the \( K \) sources, \( x_1, \ldots, x_K \) have equatorial coordinates (Right Ascension \( \alpha \), Declination \( \delta \)) equal to \((\alpha_1, \delta_1, \ldots, (\alpha_K, \delta_K)\). The aim is to find the optimum \( Q \) clusters so that the objective function \( f = \sum_{q=1}^{Q} D(L_q) \) is minimized. \( D(L_q) \) is the angular diameter of cluster \( L_q \), for \( q \in \{1, \ldots, Q\} \), defined as

\[
D(L_q) = \max \{d(x_i, x_j)|x_i, x_j \in L_q\},
\]  
(7)

and \( d(\cdot, \cdot) \) is the angular separation between any two points on the celestial sphere. Having two radio sources \( a \) and \( b \) with equatorial coordinates \((\alpha_a, \delta_a)\) and \((\alpha_b, \delta_b)\), respectively, the angular separation \( d(a, b) \), in radians, is obtained by

\[
tan^{-1}\sqrt{\cos^2\delta_b \sin^2\Delta \alpha + [\cos \delta_b \sin \delta_a - \sin \delta_b \cos \delta_a \cos \Delta \alpha]^2} - \sin \delta_a \sin \delta_b + \cos \delta_a \cos \delta_b \cos \Delta \alpha = \Delta \alpha = \alpha_b - \alpha_a.
\]  
(8)

To get the most information from the strongest observed signals in calibration, the centroids of the clusters should lean towards the brightest sources. Therefore, for defining the centroids, we associate a weight to the \( i \)-th source, \( i \in \{1, \ldots, K\} \), as

\[
w_i = w(x_i) \equiv \frac{I_i}{I^*},
\]  
(9)

where \( I_i \) is the source’s intensity and \( I^* = \min \{I_1, \ldots, I_K\} \).

We cluster radio sources using weighted K-means and divisive hierarchical clustering algorithms. Since the source clustering for calibration is performed offline, its computational complexity is negligible compared with the calibration procedure itself. Both of the clustering methods are hard clustering techniques which divide data to distinct clusters. However, we expect more accurate results using fuzzy (soft) clustering, which constructs overlapping clusters with uncertain boundaries. Application and performance of this type of clustering will be explored in future work.
3.1. Weighted K-means clustering algorithm

Step1. Select the \( Q \) brightest sources, \( x_{1^*}, \ldots, x_{Q^*} \), and initialize the centroids of \( Q \) clusters by their locations as

\[
c_q = [\alpha_{q^*}, \delta_{q^*}], \quad \text{for } q \in \{1, \ldots, Q\}, \quad q^* \in \{1^*, \ldots, Q^*\}. \tag{10}
\]

Step2. Assign each source to the cluster with the closest centroid, defining the membership function

\[
m_{L_q}(x_i) = \begin{cases} 1, & \text{if } d(x_i, c_q) = \min\{d(x_i, c_j) | j = 1, \ldots, Q\} \\ 0, & \text{otherwise} \end{cases}
\]

Step3. Update the centroids by

\[
c_q = \frac{\sum_{i=1}^{K} m_{L_q}(x_i) w_i x_i}{\sum_{i=1}^{K} m_{L_q}(x_i) w_i}, \quad \text{for } q \in \{1, \ldots, Q\}. \tag{11}
\]

Repeat steps 2 and 3 until there are no reassignments of sources to clusters.

3.2. Divisive hierarchical clustering algorithm

Step1. Initialize the cluster counter \( Q' \) to 1, assign all the \( K \) sources to a single cluster \( L_1 \) and \( \emptyset \) to a set of null clusters \( A \).

Step2. Choose cluster \( L_{q^*} \), for \( q^* \in \{1, \ldots, Q'\} - A \), with the largest angular diameter

\[
D(L_{q^*}) = \max\{D(L_q) | q \in \{1, \ldots, Q'\} - A\}. \tag{12}
\]

Step3. Apply the presented weighted K-means clustering technique to split \( L_{q^*} \) into two clusters, \( L'_{q^*} \) and \( L''_{q^*} \).

Step4. If \( D(L'_{q^*}) + D(L''_{q^*}) < D(L_{q^*}) \), then set \( Q' = Q' + 1 \), \( L_{q^*} = L'_{q^*} \), \( L_{Q'} = L''_{q^*} \), and \( A = \emptyset \); otherwise set \( A = A \cup \{q^*\} \).

Repeat steps 2, 3, and 4 until \( Q' = Q \).

4. ILLUSTRATIVE EXAMPLE

We consider the calibration of data obtained by LOFAR using 25 stations (receivers). The observation is centered at the radio source 3C196 and has an integration time of 6 hours. The central as well as the four brightest sources were initially subtracted and the result is shown in Fig. 1. We subsequently processed the same data using the classical calibration, in the direction of the 8 bright sources, and the aforementioned clustered calibration, with 10 clusters produced by weighted K-means and divisive hierarchical clustering methods, on 30 seconds time intervals. For all the methods the LS optimization is used with 9 iterations and the residual images, zoomed into the area enclosed by the white window in Fig. 1 and shown in Fig. 2. The inset figures focusing on one of those 8 sources, randomly chosen, show that the source has been considerably better subtracted in the case of clustered calibration, while in the case of classical calibration there is a significant residual error remaining. Table 1 presents the comparison of the Root Mean Squared (RMS) of the residual maps for different regions and the full images produced by the classical and clustered calibration methods. The clustered calibration method has a lower RMS in all the cases.

| RMS (mJy) | A  | B  | C  | D  | Full image |
|-----------|----|----|----|----|------------|
| Classical | 6.5| 6.0| 5.6| 6.5| 6.8        |
| HC        | 5.2| 5.3| 5.4| 5.5| 5.4        |
| WKC       | 5.5| 5.4| 5.5| 5.3| 5.9        |

Table 1. The RMS of the residual images for the LS calibration without clustering (classical calibration), and by using Hierarchical (HC) as well as Weighted K-means Clustering (WKC) of the sources. The letters A, B, C and D correspond to the regions demarcated by the boxes in Fig. 1.

5. CONCLUSIONS

We have introduced a clustered calibration scheme for calibrating radio interferometric data towards the sensitivity limit. The method upgrades the coherencies of individual sources by their total amount obtained at each source cluster. Then, it applies calibration to these new coherencies that carry a higher level of information compared with the initial
ones. Therefore, for calibration of sources below the noise level it has a considerably better performance compared with un-clustered calibration techniques. Divisive Hierarchical as well as Weighted K-means clustering methods are used to exploit the spatial proximity of the sources. It is also shown by an illustrative example that the RMS at different regions of the clustered calibration’s residual images is consistently lower, when compared to the un-clustered calibration, which reveals its superiority at a low SNR. Hierarchical clustering provides a marginally better result since it constructs clusters of smaller angular diameters and thus it assigns the same calibration solutions to sources that have smaller angular separations. Future work will address the estimation of the optimum number of clusters, the performance of fuzzy clustering in the clustered calibration, and combination of clustering with different calibration methods.

6. REFERENCES

[1] T. J. Pearson and A. C. S. Readhead, “Image Formation by Self-Calibration in Radio Astronomy,” ARA&A, vol. 22, pp. 97–130, September 1984.

[2] G. Bernardi, D. A. Mitchell, S. M. Ord, L. J. Greenhill, B. Pindor, R. B. Wayth, and J. S. B. Wyithe, “Subtraction of point sources from interferometric radio images through an algebraic forward modeling scheme,” ArXiv e-prints, December 2010.

[3] A. Liu, M. Tegmark, and M. Zaldarriaga, “Will point sources spoil 21-cm tomography?,” MNRAS, vol. 394, pp. 1575–1587, April 2009.

[4] B. Pindor, J. S. B. Wytche, D. A. Mitchell, S. M. Ord, R. B. Wayth, and L. J. Greenhill, “Subtraction of Bright Point Sources from Synthesis Images of the Epoch of Reionization,” PASA, vol. 28, pp. 46–57, July 2010.

[5] R. Xu and D. Wunsch, Clustering, John Wiley and Sons, October 2008.

[6] J. B. MacQueen, “Some methods for classification and analysis of multivariate observations,” in the fifth Berkeley Symposium on Mathematical Statistics and Probability, 1967, vol. 1, pp. 281–297.

[7] K. Kerdprasop, N. Kerdprasop, and P. Sattayatham, “Weighted k-means for density-biased clustering,” in Data Warehousing and Knowledge Discovery, 2005, pp. 488–497.

[8] S. C. Johnson, “Hierarchical clustering schemes,” Psychometrika, vol. 32, no. 3, pp. 241–254, September 1967.

[9] A. R. Thompson, J. M. Moran, and G. W. Swenson, Jr., Interferometry and Synthesis in Radio Astronomy, John Wiley & Sons, May 2001.

[10] J. P. Hamaker, J. D. Bregman, and R. J. Sault, “Understanding radio polarimetry. I. Mathematical foundations,” A&A, vol. 117, pp. 137–147, May 1996.

[11] J. P. Hamaker, “Understanding radio polarimetry. V. Making matrix self-calibration work: processing of a simulated observation,” A&A, vol. 456, pp. 395–404, September 2006.

[12] P. Lampropoulos, The LOFAR Epoch of Reionization Data Model: Simulations, Calibration, Inversion, Ph.D. thesis, Kapteyn Astronomical Institute, September 2010.

[13] S. Yatawatta, S. Zaroubi, G. de Bruyn, L. Koopmans, and J. Noordam, “Radio interferometric calibration using the SAGE algorithm,” in IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009, pp. 150 –155.

[14] S. Kazemi, S. Yatawatta, S. Zaroubi, A. G. de Bruyn, L. V. E. Koopmans, and J. Noordam, “Radio Interferometric Calibration Using The SAGE Algorithm,” MNRAS, in press.

[15] M. Born and E. Wolf, Principles of Optics, Cambridge University Press, October 1999.