Contributions to robust calibration methods in radio astronomy

Virginie Ollier

université PARIS-SACLAY

NNT : 2018SACLN023

Thèse de doctorat

école normale supérieure paris-saclay

Contributions aux méthodes de calibration robuste en radioastronomie

Contributions to robust calibration methods in radio astronomy

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'École Normale Supérieure

École doctorale n°580 Sciences et Technologies de l'Information et de la Communication (STIC)
Spécialité de doctorat : Traitement du Signal et des Images

Thèse présentée et soutenue à Cachan, le 5 Juillet 2018, par

Virginie Ollier

Composition du Jury :

M. Christian Jutten Professeur, Université Grenoble-Alpes (GIPSA-lab)	Président
M. David Brie Professeur, Université de Lorraine (CRAN)	Rapporteur
M. Jean-François Giovannelli Professeur, Université de Bordeaux (IMS)	Rapporteur
Mme. Sylvie Marcos Directrice de recherche, CNRS (L2S)	Examinatrice
M. Jean-Luc Starck Directeur de recherche, CEA Paris-Saclay (CosmoStat)	Examinateur
M. Pascal Larzabal Professeur, Université Paris-Sud (SATIE)	Directeur de thèse
M. Rémy Boyer Maître de conférences, Université Paris-Sud (L2S)	Co-Encadrant
M. Mohammed Nabil El Korso Maître de conférences, Université Paris-Nanterre (LEME)	Co-Encadrant
M. André Ferrari Professeur, Université de Nice Sophia Antipolis (J.L. Lagrange)	Invité

To my loved ones

Remerciements

Une thèse, c'est la chance de pouvoir travailler pendant trois ans sur un sujet qui nous tient à coeur et acquérir ainsi une expertise technique qui constitue un atout essentiel d'un point de vue professionnel. Mais c'est aussi des échanges et des rencontres avec une multitude de personnes qui font de la thèse une véritable aventure humaine.

Je tiens tout d'abord à remercier l'ensemble des membres du jury pour l'intérêt qu'ils ont porté à mon travail et leurs remarques qui ont permis de l'enrichir. Je remercie David Brie et Jean-François Giovannelli d'avoir accepté d'être rapporteurs sur ma thèse et qui m'ont apporté des commentaires très constructifs et utiles. Merci à Sylvie Marcos et Jean-Luc Starck d'avoir accepté d'évaluer mes travaux lors du suivi à mi-parcours, avec Jean-François Giovannelli également, et qui étaient aussi présents lors de ma soutenance en tant qu'examinateurs, avec des remarques toujours pertinentes. Je remercie Christian Jutten de m'avoir fait l'honneur de présider ce comité et d'apporter son avis sur ce travail.

Selon moi, la qualité d'une thèse se mesure essentiellement par l'encadrement. Et de ce point de vue là, j'ai été très chanceuse.

Merci Pascal pour tous les bons conseils que tu as pu me donner, pour ton enthousiasme sans faille, ton expérience sur un très grand nombre de sujets et ton soutien qui m'a permis de garder confiance tout le long de la thèse. Merci aussi de m'avoir fait confiance et de m'avoir donné l'opportunité de faire cette thèse après le stage, en me laissant aussi la chance de voyager vers des destinations incroyables lors de conférences ou écoles thématiques. Merci pour tout.

Merci Rémy pour ton aide très précieuse, ta très forte capacité de travail et ta très grande disponibilité. J'ai aussi eu la chance de pouvoir effectuer mes tout premiers enseignements à tes côtés et ce fut une expérience très agréable. Tu es toujours de très bon conseil avec une très grande rigueur scientifique et j'espère vraiment que tu te plairas à Lille !

Merci Nabil. Tu as toujours été très présent, que ce soit pendant le stage ou la thèse, très disponible, investi et à l'écoute. Tu as pu m'aider au jour le jour, répondre à toutes les questions que je pouvais avoir, me donner de nombreux conseils et m'accorder beaucoup de temps avec une patience et un soutien indéfectibles. Merci pour cela, j'ai vraiment apprécié pouvoir travailler à tes côtés.

Grâce à vous, cette thèse fut pour moi une expérience inoubliable et j'espère que nous pourrons toujours garder contact et nous revoir dans les années à venir.

Une thèse, c'est aussi un environnement de recherche privilégié au sein de laboratoires : le Satie et le L2S dans mon cas. Je tiens à remercier l'ensemble des personnes que j'ai pu côtoyer de près ou de loin et qui rendent l'atmosphère de travail toujours très chaleureuse et agréable, de par leur sympathie et leur bonne humeur. Je pense tout particulièrement à Eric V., Jean-Pierre, Gwenaël, Hélène, Sophie, Béatrice, Aurore, Dorothée, Fabien, Dominique, David, Muriel, Yohan, Mehdi, Alexiane, Sakina mais aussi Alex, Lucien, Gordana, Yassine, Arnaud, Rayen, Jean-Phi, Chengfang, Bruno, Ammar, Eugénie, Eric C., Rita, Joana, Elisa, Abigael, Martin, Alice, Cyrile,

Abdellatif, Guillaume, Maxime, Julien et tant d'autres... J'aimerais particulièrement remercier André Ferrari pour son accueil à Nice et toute l'aide qu'il m'a apportée afin de prendre en main le logiciel MeqTrees. Son expertise fut clairement indispensable et d'un grand atout pour ce travail. Merci d'avoir accepté l'invitation à ma soutenance de thèse et pour tous les échanges que nous avons eus, qui ont donné lieu à plusieurs publications scientifiques ensemble. J'en profite également pour remercier l'ANR MAGELLAN qui a permis à cette thèse de voir le jour, ainsi que l'ensemble des membres qui y ont participé dont Chiara Ferrari, David Mary, Rémi Flamary, Pascal Bianchi,... Also many thanks to Marius Pesavento for his help and his comments on papers we wrote together. I really appreciated meeting you in Cachan just before the beginning of my PhD and also at conferences.

Une thèse, c'est aussi la possibilité de s'ouvrir intellectuellement et faire une multitude de choses différentes. Dans mon cas, j'ai eu la chance d'effectuer une mission d'enseignement au sein de l'IUT de Cachan. Merci Nicolas de m'avoir accueilli chaleureusement au sein du département GEII1. Merci à Pierre-Yves, Sabine, Yannick, Philippe, Nhung, Marie et un grand merci à Stéphane pour toutes ces heures de TP que nous avons passées ensemble. Le temps que tu as passé avec moi lors des préparations et les discussions que nous avons pu avoir ensemble m'ont convaincue que tu étais un enseignant passionnant et passionné, très impliqué pour ses élèves.

Une thèse, c'est enfin savoir rebondir et toujours rester positif. Le soutien de mes proches, amis et famille, fut essentiel et je ne les remercierai jamais assez pour cela, en particulier mes parents, mon frère et ma soeur. Merci à M. Ulrich d'avoir toujours eu confiance en mes capacités et de m'avoir donné le goût des mathématiques. Pour finir, merci à toi, Thomas, d'avoir toujours été présent à mes côtés et de m'avoir épaulée quelle que soit la situation.

Virginie Ollier
Juillet 2018

Satisfaction lies in the effort, not in the attainment, full effort is full victory.
Mahatma Gandhi

Abstract

In radio astronomy, the design of observation instruments has evolved significantly through the years. Instead of considering a single large disk to probe the sky, the development of interferometric systems has enabled to provide higher resolution images. Indeed, the signals captured by each separate radio telescope are combined through synthesis imaging: the resulting angular resolution is the same as for a single large-aperture telescope. Nowadays, such phased arrays are widely used: the beam is electronically steered in different directions without physically moving the antennas, leading to more flexibility. The new generation of radio telescopes is based on very large arrays as the Low Frequency Array (LOFAR) in Europe with about 50 000 antennas and the future Square Kilometre Array (SKA) in Australia and South Africa which will count 2000 high and mid-frequency dishes and aperture arrays and a million low-frequency antennas. These are the most sensitive radio instruments and they will provide fundamental new studies about our galaxy. However, the huge number of antennas in the array, the large collecting area and dynamic range impose new challenges to overcome. To meet the theoretical optimal performances of such ambitious international projects, signal processing issues in terms of calibration, data reduction and imaging need to be addressed.

Accurate calibration is of critical importance in order to recover high resolution images with no distortions. This process consists in correcting for all environmental and instrumental effects which corrupt the observations and it reveals to be a daunting parameter estimation task as direction dependent perturbations are prevailing. Thus, developing estimation algorithms with reasonable computational cost is essential. Most state-of-the-art calibration approaches are least-squares based, using known reference sources as calibration sources thanks to tables describing accurately their position and flux. They are theoretically asymptotically efficient under a Gaussian noise model and operate mostly in an iterative manner for a mono-frequency scenario. However, in practice, the Gaussian classical noise assumption is not valid as radio frequency interference affects the measurements, multiple unknown weak sources appear within the wide field-of-view and transient phenomena can randomly create outliers. Furthermore, considering one frequency bin at a time with a single centralized agent processing all data leads to suboptimality and computational limitations. The goal of this thesis is to explore robustness of calibration algorithms w.r.t. the presence of outliers in a multi-frequency scenario.

To this end, we propose the use of an appropriate noise model, namely the so-called coumpound-Gaussian which encompasses a broad range of different heavy-tailed distributions, including the Student's t, K-distribution, Laplace, etc. To combine limited computational complexity and quality of calibration, we designed an iterative calibration algorithm based on the maximum likelihood estimator and the compound-Gaussian modeling. Using the Expectation Maximization (EM) and the Block Coordinate Descent (BCD) algorithms results in closed-form expressions for the estimates. In addition, a computationally efficient way to handle multiple sub-frequency bands is to apply distributed and decentralized strategies. Thus, the global operational load is distributed over a network of computational agents and calibration amounts to solve a global

constrained problem thanks to available variation models or by assuming smoothness across frequency. A particularly well-suited method for such large-scale problems is the Alternating Direction Method of Multipliers (ADMM) which we applied in our work.

Keywords: Parametric estimation, calibration, robustness, distributed optimization, radio astronomy, array processing.

Résumé

Une réelle mutation technologique s'opère dans la conception des instruments d'observation pour la radioastronomie. En effet, les nouveaux réseaux phasés à balayage électronique viennent supplanter les traditionnelles antennes paraboliques à balayage mécanique et s'appuyent sur l'interférométrie pour obtenir des images de meilleure résolution. Les signaux radio mesurés par chaque antenne sont alors corrélés entre eux. Grâce à la synthèse d'ouverture, la résolution angulaire résultante est équivalente à celle obtenue avec un unique large télescope dont le diamètre équivaudrait à la taille du réseau. Ces nouveaux réseaux d'antennes sont particulièrement flexibles puisqu'ils sont orientés électroniquement selon différentes directions possibles, sans aucun mouvement mécanique des antennes. Parmi la nouvelle génération d'instruments très sensibles, nous pouvons mentionner le LOFAR (*Low Frequency Array*) qui regroupe près de 50 000 antennes à travers l'Europe mais aussi le futur radiotélescope géant SKA (*Square Kilometre Array*) dont les stations seront réparties à terme en Australie et en Afrique du Sud et fourniront de nouvelles études essentielles sur notre galaxie. Néanmoins, la grande surface collectrice du réseau, le très large champ de vision des antennes et la forte dynamique des images nécessitent de relever un certain nombre de défis scientifiques en termes de calibration, réduction des données et reconstruction d'images.

En radioastronomie, les signaux d'intérêt sont perturbés par de nombreux effets environnementaux et instrumentaux, nécessitant la mise en oeuvre de techniques algorithmiques pour les traiter et pouvoir ainsi reconstruire *in fine* des images parfaitement nettes de l'espace. Cette étape de correction des perturbations se nomme la calibration et repose généralement sur une méthode des moindres carrés, en s'appuyant sur des sources de référence dont la position et l'intensité sont connues. Cette méthode classique se révèle être théoriquement asymptotiquement efficace dans le cas d'un bruit Gaussien et fonctionne généralement de manière itérative pour une seule fréquence considérée. Cependant, en pratique, l'hypothèse du bruit Gaussien n'est pas toujours valide car de multiples sources inconnues de faible intensité sont visibles dans le champ de vision et des interférences radioélectriques perturbent les données pour finalement modifier la nature typiquement gaussienne du bruit. En outre, réaliser une calibration fréquence par fréquence, de façon complètement indépendante, n'est pas la manière la plus optimale de procéder. Le but de ce travail est donc de développer des algorithmes de correction dans le traitement des signaux radio qui soient robustes à la présence d'éventuelles valeurs aberrantes ou sources d'interférences, et qui soient adaptés au contexte multi-fréquentiel.

Par conséquent, nous nous appuyons sur une modélisation plus générale que la loi gaussienne, appelée processus Gaussien composé, qui inclut un grand nombre de distributions différentes telles que la loi de Student ou la loi de Laplace. L'algorithme robuste que nous proposons est itératif et basé sur l'estimation au sens du maximum de vraisemblance. Pour réduire les temps de calcul, nous avons recours aux algorithmes EM (*Expectation Maximization*) et BCD (*Block Coordinate Descent*) qui, pour un choix judicieux de paramétrisation, fournissent des expressions analytiques pour les paramètres d'intérêt. En accord avec le scénario multi-fréquentiel sous étude,

nous exploitons la variation spectrale des perturbations en utilisant des méthodologies telles que l'optimisation distribuée sous contraintes et le traitement parallèle des données, notamment avec l'algorithme ADMM (*Alternating Direction Method of Multipliers*).

Mots clés : Estimation paramétrique, calibration, robustesse, optimisation distribuée, radioastronomie, traitement d'antennes.

Contents

List of Figures

List of Tables

Notations

Acronyms

- ADI: Alternating Direction Implicit.
- ADMM: Alternating Direction Method of Multipliers.
- AIPS: Astronomical Image Processing System.
- ASKAP: Australian Square Kilometre Array Pathfinder.
- ASTRON: Netherlands Institute for Radio Astronomy.
- BBS: Black-Board Self-cal.
- BCD: Block Coordinate Descent.
- BFGS: Broyden-Fletcher-Goldfarb-Shanno.
- CASA: Common Astronomy Software Applications.
- CasA: Cassiopeia A.
- CG: Compound-Gaussian.
- COLR: Column Ratio.
- CRB: Cramer-Rao Bound.
- CS: Cotton-Schwab.
- CygA: Cygnus A.
- DEC: Declination.
- DOA: Direction Of Arrival.
- ECME: Expectation-Conditional Maximization Either.
- EM: Expectation Maximization.
- FBC: Field-Based Calibration.
- FIM: Fisher Information Matrix.
- GNLS: Generalized Nonlinear Least Squares.
- GPU: Graphics Processing Unit.
- IG-CG: Inverse-Gaussian Compound-Gaussian.
- i.i.d.: independent and identically distributed.
- IMAPE: Iterative MAP Estimator.

- IRLS: Iterative Reweighted Least Squares.
- LM: Levenberg-Marquardt.
- LOFAR: Low Frequency Array.
- LOGLS: Logarithmic Least Squares.
- LS: Least-Squares.
- LS-MVI: Least Squares Minimum Variance Imaging.
- MAP: Maximum A Posteriori.
- MeerKAT: Karoo Array Telescope.
- MEM: Maximum Entropy Method.
- ML: Maximum Likelihood.
- MMSE: Minimum Mean Squared Error.
- MRCA: Multi-frequency Robust Calibration Algorithm.
- MRCA^{R3}: Multi-frequency Robust Calibration Algorithm for Regime 3.
- MS: Measurement Set.
- MSE: Mean Square Error.
- MWA: Murchison Widefield Array.
- OS: Ordered Subsets.
- pdf: probability density function.
- PSF: Point Spread Function.
- RA: Right Ascension.
- RCMLE: Relaxed Concentrated ML Estimator.
- RM: Rotation Measure.
- RMT: Random Matrix Theory.
- RTR: Riemannian Trust-Region.
- SA: Simulated Annealing.
- SAGE: Space Alternating Generalized Expectation Maximization.
- SKA: Square Kilometre Array.
- SNR: Signal-to-Noise Ratio.
- SPAM: Source Peeling and Atmospheric Modeling.
- StEFCal: Statistically EFficient and Fast Calibration.
- SUMSS: Sydney University Molonglo Sky Survey.
- TEC: Total Electron Content.
- VLA: Very Large Array.
- (W)ALS: (Weighted) Alternating Least-Squares.
- WSF: Weighted Subspace Fitting.
- WSRT: Westerbork Synthesis Radio Telescope.

Mathematical symbols

- $\mathbb{C}$: complex numbers.
- $\mathbb{R}$: real numbers.
- $\Re\{\cdot\}$: real part of a complex number.
- $\mathrm{Im}\{\cdot\}$: imaginary part of a complex number.
- a : scalar quantity.
- $\mathbf{a}$: column vector.
- $\mathbf{A}$: matrix.
- $(\cdot)^*$: complex conjugate operator.
- $(\cdot)^T$: transpose operator.
- $(\cdot)^H$: Hermitian transpose operator.
- $\mathbf{I}_N$: $N \times N$ identity matrix.
- $|\cdot|$: determinant operator.
- $||\cdot||_2$: l_2 norm.
- $||\cdot||_F$: Frobenius norm.
- $[\mathbf{a}]_k$: k-*th* entry of vector $\mathbf{a}$.
- $[\mathbf{A}]_{i,j}$: entry in the i-*th* row and j-*th* column of matrix $\mathbf{A}$.
- $[\mathbf{A}]_{:,k}$: k-*th* column vector of matrix $\mathbf{A}$.
- $\mathrm{tr}\{\cdot\}$: trace operator.
- $\mathrm{vec}(\cdot)$: vectorization operator to stack all columns of a matrix on top of one another.
- $\mathcal{N}(\mathbf{m}, \mathbf{C})$: Gaussian probability density function with mean $\mathbf{m}$ and covariance matrix $\mathbf{C}$.
- $p(x)$: probability density function.
- $p(x, y)$: joint probability density function.
- $p(x|\, y)$: conditional probability density function.
- $\mathrm{E}\{\cdot\}$: expectation operator.
- $\otimes$: Kronecker product between two matrices. For two coordinate vectors, we refer to it as the outer product.
- bdiag$\{\cdot\}$: block-diagonal operator.
- diag$\{\cdot\}$: conversion of a vector into a diagonal matrix.
- j : complex number whose square equals -1.

Résumé étendu

Introduction

Le travail présenté dans ce document porte sur l'étude des nouveaux radio-interféromètres tels que le LOFAR (*Low Frequency Array*) [7] en Europe ou le futur SKA (*Square Kilometre Array*) [8] dont les stations seront déployées en Australie et en Afrique du Sud. En tant que réel instrument d'envergure, le SKA sera constitué de millions d'antennes formant l'équivalent d'une surface de réception d'un kilomètre carré et capables de capter des ondes radio sur un large spectre de fréquences. En sondant très rapidement le ciel avec une précision sans précédent, ce système permettra d'obtenir une qualité d'image supérieure aux installations actuelles et complètera nos connaissances sur les lois physiques régissant l'univers. Néanmoins, un tel potentiel nécessite de relever un certain nombre de défis scientifiques et technologiques [9–11]. L'imagerie du ciel requiert notamment une estimation précise des différentes perturbations introduites le long du trajet du signal, qu'elles soient d'origine environnementale ou instrumentale, et ce, afin de reconstruire des images haute résolution, sans déformations.

Dans ce document, nous nous intéressons plus particulièrement à l'étape de calibration qui consiste à corriger les perturbations subies par le signal radio depuis son émission par une source jusqu'à sa réception par le réseau d'antennes.

Parmi les effets perturbateurs, nous pouvons citer

- la présence des gains électroniques complexes des antennes.
- l'anisotropie dans la réponse des antennes (réponse non uniforme, d'où une dépendance selon la direction d'arrivée).
- le retard de propagation dû au passage du signal par la troposphère et l'ionosphère. Ces couches spécifiques de l'atmosphère se caractérisent par des phénomènes de diffraction et de réfraction, variables dans le temps et l'espace, et affectent notamment la polarisation du signal.

En calibration, la plupart des algorithmes s'appuient sur une méthode de type moindres carrés, en minimisant une fonction coût par rapport à chaque paramètre inconnu, de façon alternée [12, 13]. Afin de réduire le coût calculatoire, des algorithmes tels que l'EM (*Expectation Maximization*) ou le SAGE (*Space Alternating Generalized Expectation Maximization*) ont déjà été proposés [14]. Malheureusement, l'hypothèse de bruit Gaussien est souvent considéré et n'est généralement pas vérifié en pratique. Cette non-gaussianité du bruit s'explique par la présence de sources inconnues de très faible intensité ou des interférences causées par des évènements ponctuels ou encore des brouillages artificiels, créant ainsi aléatoirement des valeurs aberrantes dans les données mesurées [15]. À ce jour et à notre connaissance, la seule tentative de calibration robuste repose sur une modélisation bien spécifique du bruit : la loi de Student, avec des variables

indépendantes et identiquement distribuées [2]. Nous proposons ici d'avoir recours à une plus large gamme de distributions, qui inclut entre autres la loi de Student et le cas classique Gaussien. Plus particulièrement, nous nous intéressons au processus dit Gaussien composé, qui s'écrit comme le produit d'une variable aléatoire positive, appelée texture, et d'un processus aléatoire Gaussien, le speckle [16, 17]. À partir de ce modèle, nous développons la méthode du maximum de vraisemblance pour estimer à la fois les paramètres de nuisance (texture et speckle) ainsi que les paramètres d'intérêt (matrices de Jones, introduites ci-après). Dans un premier temps, nous nous focalisons sur le cas mono-fréquence.

Considérons D sources de calibration et M antennes dans le réseau de capteurs. Chaque signal incident se décompose selon deux axes de polarisation orthogonales (x, y). Le signal émis par la i-ème source $\mathbf{e}_i$ et la tension mesurée par la p-ème antenne $\mathbf{v}_{i,p}(\boldsymbol{\theta})$ sont liés selon la relation suivante [18, 19]

$$\mathbf{v}_{i,p}(\boldsymbol{\theta}) = \mathbf{J}_{i,p}(\boldsymbol{\theta})\mathbf{e}_i \tag{1}$$

où $\mathbf{J}_{i,p}(\boldsymbol{\theta})$ est une matrice de Jones de taille 2×2, paramétrée par le vecteur inconnu $\boldsymbol{\theta}$. Elle tient compte de toutes les perturbations introduites le long du trajet depuis la source i jusqu'à l'antenne p. Puisqu'il existe une matrice de Jones différente pour chaque chemin i-p, le nombre total de matrices à estimer est donc DM.

Le principe d'un interféromètre est de mesurer les corrélations des signaux récoltés par chaque paire d'antennes (p, q). Dans le cas non bruité, cette corrélation s'écrit comme suit

$$\mathbf{V}_{pq}(\boldsymbol{\theta}) = \mathrm{E}\Big\{ \sum_{i=1}^{D} \mathbf{v}_{i,p}(\boldsymbol{\theta})\mathbf{v}_{i,q}^H(\boldsymbol{\theta}) \Big\} = \sum_{i=1}^{D} \mathbf{J}_{i,p}(\boldsymbol{\theta})\mathbf{C}_i\mathbf{J}_{i,q}^H(\boldsymbol{\theta}) \quad \text{avec} \quad p < q, \quad p, q \in \{1, \ldots, M\}^2, \tag{2}$$

où $\mathbf{C}_i = \mathrm{E}\{\mathbf{e}_i\mathbf{e}_i^H\}$ décrit l'état de polarisation de la i-ème source, supposé connu, si il s'agit d'une source de calibration. De manière équivalente, nous pouvons considérer

$$\tilde{\mathbf{v}}_{pq}(\boldsymbol{\theta}) = \mathrm{vec}\Big(\mathbf{V}_{pq}(\boldsymbol{\theta})\Big) = \sum_{i=1}^{D} \mathbf{s}_{i,pq}(\boldsymbol{\theta}) \tag{3}$$

où $\mathbf{s}_{i,pq}(\boldsymbol{\theta}) = \Big(\mathbf{J}_{i,q}^*(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,p}(\boldsymbol{\theta})\Big)\, \mathbf{c}_i$ et $\mathbf{c}_i = \mathrm{vec}(\mathbf{C}_i)$.

En pratique, les données sont perturbées par un vecteur bruit $\mathbf{n}_{pq}$ qui prend en compte le bruit ambiant Gaussien mais aussi la présence d'éventuelles valeurs aberrantes. Ces dernières peuvent être dues à des phénomènes d'interférences ou encore à la présence de sources de très faible intensité qui nous sont inconnues et inaccessibles dans un premier temps.

Finalement, le vecteur de corrélation bruité de taille 4×1 s'écrit

$$\mathbf{v}_{pq} = \tilde{\mathbf{v}}_{pq}(\boldsymbol{\theta}) + \mathbf{n}_{pq}. \tag{4}$$

En notant $B = \frac{M(M-1)}{2}$ le nombre total de paires d'antennes (p, q), le vecteur global de données $\mathbf{y} \in \mathbb{C}^{4B \times 1}$, regroupant toutes les corrélations mesurées par l'interféromètre, est donné par

$$\mathbf{y} = [\mathbf{v}_{12}^T, \mathbf{v}_{13}^T, \ldots, \mathbf{v}_{(M-1)M}^T]^T = \sum_{i=1}^{D} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n} \tag{5}$$

tel que $\mathbf{s}_i(\boldsymbol{\theta}) = \Big[\mathbf{s}_{i,12}^T(\boldsymbol{\theta}), \mathbf{s}_{i,13}^T(\boldsymbol{\theta}), \ldots, \mathbf{s}_{i,(M-1)M}^T(\boldsymbol{\theta})\Big]^T$ et $\mathbf{n} = \Big[\mathbf{n}_{12}^T, \mathbf{n}_{13}^T, \ldots, \mathbf{n}_{(M-1)M}^T\Big]^T$.

Méthode de calibration robuste

En radioastronomie, le bruit ne peut tout simplement pas être considéré comme Gaussien du fait de la présence d'interférences et de sources d'intérêt de très faible intensité, autres que les D sources de calibration supposées connues dans le modèle du ciel. Nous traitons ici la notion de robustesse à l'aide d'un modèle Gaussien composé pour le bruit, défini comme suit pour chaque paire d'antennes

$$\mathbf{n}_{pq} = \sqrt{\tau_{pq}}\ \mathbf{g}_{pq}, \tag{6}$$

où le facteur de puissance réel τ_{pq} est positif et aléatoire alors que le vecteur $\mathbf{g}_{pq}$ suit une loi gaussienne de moyenne nulle, i.e., $\mathbf{g}_{pq} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega})$. Pour éviter tout problème d'ambiguïté, une contrainte, purement arbitraire, est requise sur la matrice de covariance $\mathbf{\Omega}$. Dans ce travail, nous supposerons que $\mathrm{tr}\{\mathbf{\Omega}\} = 1$.

En supposant des mesures indépendantes entre elles, la vraisemblance admet l'expression suivante

$$p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{\Omega}) = \prod_{pq} \frac{1}{|\pi \tau_{pq} \mathbf{\Omega}|} \exp\left\{ -\frac{1}{\tau_{pq}} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \mathbf{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}) \right\}, \tag{7}$$

où le vecteur $\boldsymbol{\tau} = [\tau_{12}, \tau_{13}, \ldots, \tau_{(M-1)M}]^T$ regroupe les réalisations de la texture et $\mathbf{u}_{pq}(\boldsymbol{\theta}) = \mathbf{v}_{pq} - \tilde{\mathbf{v}}_{pq}(\boldsymbol{\theta})$. Ainsi, la log-vraisemblance s'écrit

$$\ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{\Omega}) = -4B \ln \pi - 4 \sum_{pq} \ln \tau_{pq} - B \ln |\mathbf{\Omega}| - \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \mathbf{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}). \tag{8}$$

Le principe de la méthode proposée consiste à estimer séquentiellement les paramètres $\boldsymbol{\theta}$, $\boldsymbol{\tau}$ et $\mathbf{\Omega}$. Ces estimations sont obtenues en maximisant la log-vraisemblance (8) par rapport à chaque paramètre inconnu, les autres paramètres étant supposés fixés [20]. Nous remarquons qu'il est nécessaire de spécifier la densité de probabilité du paramètre texture τ_{pq} dans (6). Pour cela, nous pouvons considérer différentes distributions *a priori* telles que les lois Gamma, inverse-Gamma, exponentielle et inverse-gaussienne, qui fournissent des formes spécifiques pour le modèle Gaussien composé et permettent de dériver différents estimateurs basés sur le MAP (*Maximum A Posteriori*) [21]. Cependant, supposer des paramètres τ_{pq} inconnus et déterministes permet d'assurer plus de flexibilité et de robustesse [22, 23]. Avec cette hypothèse et la méthodologie itérative adoptée, nous obtenons les estimations des paramètres inconnus décrites ci-dessous

1) Estimation de τ_{pq} : Annuler la dérivée de (8) par rapport à τ_{pq} conduit à

$$\hat{\tau}_{pq} = \frac{1}{4} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \mathbf{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}). \tag{9}$$

2) Estimation de $\mathbf{\Omega}$: Par une méthodologie similaire, nous obtenons l'expression suivante pour la partie speckle

$$\hat{\mathbf{\Omega}}^{t+1} = \frac{4}{B} \sum_{pq} \frac{\mathbf{u}_{pq}(\boldsymbol{\theta}) \mathbf{u}_{pq}^H(\boldsymbol{\theta})}{\mathbf{u}_{pq}^H(\boldsymbol{\theta}) \left(\hat{\mathbf{\Omega}}^t \right)^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta})} \tag{10}$$

où t fait référence à la t-ème itération. Finalement, la contrainte spécifiée préalablement sur la matrice de speckle conduit à la normalisation suivante

$$\hat{\mathbf{\Omega}}^{t+1} = \frac{\hat{\mathbf{\Omega}}^{t+1}}{\mathrm{tr}\left\{ \hat{\mathbf{\Omega}}^{t+1} \right\}}. \tag{11}$$

3) Estimation de $\boldsymbol{\theta}$: Conditionnellement à $\hat{\boldsymbol{\tau}}$ et $\hat{\boldsymbol{\Omega}}$, l'estimation de $\boldsymbol{\theta}$ revient à effectuer l'opération suivante

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \left\{ \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}) \right\}. \tag{12}$$

Nous distinguons alors deux cas de figures : dans le premier, les matrices de Jones sont dites non structurées et nous cherchons alors à estimer les quatre éléments complexes de chaque matrice $\mathbf{J}_{i,p}(\boldsymbol{\theta})$ pour $i \in \{1, \ldots, D\}$ et $p \in \{1, \ldots, M\}$. Dans l'autre cas dit structuré, chaque matrice de Jones admet une décomposition spécifique selon les perturbations physiques considérées et $\boldsymbol{\theta}$ fait référence aux éléments introduits dans le modèle paramétrique [24–26].

Matrices de Jones non structurées

Pour réaliser la minimisation en (12), des méthodes avec un coût de calcul réduit sont requises. En effet, le vecteur d'intérêt $\boldsymbol{\theta}$ contient un très grand nombre d'éléments. Notons que ce dernier peut se décomposer comme suit

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_D^T]^T = [\boldsymbol{\theta}_{1,1}^T, \ldots, \boldsymbol{\theta}_{1,M}^T, \ldots, \boldsymbol{\theta}_{D,1}^T, \ldots, \boldsymbol{\theta}_{D,M}^T]^T, \tag{13}$$

où $\boldsymbol{\theta}_{i,p} \in \mathbb{R}^{8\times 1}$ décrit le trajet *i-p*. Ainsi, il est judicieux d'écrire $\mathbf{J}_{i,p}(\boldsymbol{\theta}) \Rightarrow \mathbf{J}_{i,p}(\boldsymbol{\theta}_{i,p})$.

Pour réduire les temps de calcul, nous avons recours à l'algorithme EM qui calcule les estimations des paramètres d'intérêt à l'aide de deux étapes successives : les étapes E et M, et tend asymptotiquement vers l'estimation au sens du maximum de vraisemblance [27,28]. Dans notre cas, l'étape E se résume à calculer l'espérance conditionnelle de données dites complètes, sachant les données observées et les paramètres estimés. Ensuite, l'étape M maximise la vraisemblance conditionnelle, soit de manière numérique avec l'algorithme LM (*Levenberg-Marquardt*) par exemple [29], soit de manière analytique si possible. Ainsi, (12) peut se décomposer en plusieurs sous-problèmes de plus faible complexité : l'optimisation se fait par rapport à $\boldsymbol{\theta}_i \in \mathbb{C}^{4M\times 1}$ au lieu de $\boldsymbol{\theta} \in \mathbb{C}^{4DM\times 1}$. Les deux étapes successives de l'EM s'écrivent comme suit

1) Étape E : Soit le vecteur de données complètes $\mathbf{w} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_D^T]^T$, où

$$\mathbf{w}_i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n}_i \tag{14}$$

tel que $\mathbf{y} = \sum_{i=1}^D \mathbf{w}_i$ et $\mathbf{n} = \sum_{i=1}^D \mathbf{n}_i$. Les vecteurs bruit $\mathbf{n}_i$ sont supposés statistiquement indépendants et générés selon $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \beta_i \boldsymbol{\Psi})$ avec $\sum_{i=1}^D \beta_i = 1$ et

$$\boldsymbol{\Psi} = \operatorname{bdiag}\Big\{ \tau_{12}\boldsymbol{\Omega}, \ldots, \tau_{(M-1)M}\boldsymbol{\Omega} \Big\}. \tag{15}$$

L'espérance conditionnelle caractéristique de l'étape E de l'EM est alors donnée par

$$\hat{\mathbf{w}}_i = \mathrm{E}\Big\{ \mathbf{w}_i | \mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega} \Big\} = \mathbf{s}_i(\boldsymbol{\theta}_i) + \beta_i \left(\mathbf{y} - \sum_{l=1}^D \mathbf{s}_l(\boldsymbol{\theta}_l) \right). \tag{16}$$

2) Étape M : Une fois les $\mathbf{w}_i$ obtenus avec (16) pour $i \in \{1, \ldots, D\}$, les $\boldsymbol{\theta}_i$ sont estimés en maximisant la vraisemblance suivante

$$p(\hat{\mathbf{w}}|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega}) = \prod_{i=1}^D \frac{1}{|\pi\beta_i\boldsymbol{\Psi}|} \exp\left\{ -\Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big)^H (\beta_i\boldsymbol{\Psi})^{-1} \Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big) \right\}. \tag{17}$$

Il faut donc considérer la fonction de coût $\phi_i(\boldsymbol{\theta}_i) = \Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big)^H (\beta_i \boldsymbol{\Psi})^{-1} \Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big)$ pour la i-ème source. Auparavant, nous avons chercher à réaliser la minimisation en (12) pour chaque $\boldsymbol{\theta}_i$. De manière similaire, nous allons par la suite maximiser la vraisemblance en (17) pour chaque $\boldsymbol{\theta}_{i,p}$, de manière alternée, grâce à l'algorithme BCD (*Block Coordinate Descent*) [30] et ce, toujours dans un souci de réduction de complexité.

Nous cherchons à minimiser la fonction $\phi_i(\cdot)$ par rapport à $\boldsymbol{\theta}_{i,p} \in \mathbb{C}^{4\times 1}$, tout en fixant les autres $\boldsymbol{\theta}_{i,q}$ pour $q \neq p$. La procédure est répétée séquentiellement pour $p \in \{1, \ldots, M\}$ et fournit les résultats suivants

$$\hat{\boldsymbol{\theta}}_{i,p} = \begin{cases} (\boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \boldsymbol{\Sigma}_i + \boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \boldsymbol{\Upsilon}_i)^{-1} (\boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \mathbf{w}_{i,p} + \boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \tilde{\mathbf{w}}_{i,p}) & \text{pour } \ 1 < p < M \\ (\boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \boldsymbol{\Sigma}_i)^{-1} \boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \mathbf{w}_{i,p} & \text{pour } \ p = 1 \\ (\boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \boldsymbol{\Upsilon}_i)^{-1} \boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \tilde{\mathbf{w}}_{i,p} & \text{pour } \ p = M \end{cases} \tag{18}$$

où

- $\mathbf{w}_{i,p} = [\mathbf{w}_{i,p(p+1)}^T, \ldots, \mathbf{w}_{i,pM}^T]^T$ et $\tilde{\mathbf{w}}_{i,p} = [\mathbf{w}_{i,1p}^{*T}, \ldots, \mathbf{w}_{i,(p-1)p}^{*T}]^T$,
- $\boldsymbol{\Sigma}_i = [\boldsymbol{\Sigma}_{i,p+1}^T, \cdots, \boldsymbol{\Sigma}_{i,M}^T]^T$ et $\boldsymbol{\Upsilon}_i = [\boldsymbol{\Upsilon}_{i,1}^{*T}, \cdots, \boldsymbol{\Upsilon}_{i,p-1}^{*T}]^T$,
- $\mathbf{A}_{i,p} = \text{bdiag}\{\beta_i \tau_{p(p+1)} \boldsymbol{\Omega}, \ldots, \beta_i \tau_{pM} \boldsymbol{\Omega}\}^{-1}$ et $\tilde{\mathbf{A}}_{i,p} = \text{bdiag}\{\beta_i \tau_{1p} \boldsymbol{\Omega}^*, \ldots, \beta_i \tau_{(p-1)p} \boldsymbol{\Omega}^*\}^{-1}$.

Nous avons également utilisé les notations suivantes :

$$\boldsymbol{\Sigma}_{i,q} = \begin{bmatrix} \alpha_{i,q} & \beta_{i,q} & 0 & 0 \\ 0 & 0 & \alpha_{i,q} & \beta_{i,q} \\ \gamma_{i,q} & \rho_{i,q} & 0 & 0 \\ 0 & 0 & \gamma_{i,q} & \rho_{i,q} \end{bmatrix}, \quad \boldsymbol{\Upsilon}_{i,q} = \begin{bmatrix} \lambda_{i,q} & \mu_{i,q} & 0 & 0 \\ \nu_{i,q} & \xi_{i,q} & 0 & 0 \\ 0 & 0 & \lambda_{i,q} & \mu_{i,q} \\ 0 & 0 & \nu_{i,q} & \xi_{i,q} \end{bmatrix}$$

avec

- $\alpha_{i,q} = q_{i_1}^* c_{i_1} + q_{i_2}^* c_{i_3}$, $\beta_{i,q} = q_{i_1}^* c_{i_2} + q_{i_2}^* c_{i_4}$, $\gamma_{i,q} = q_{i_3}^* c_{i_1} + q_{i_4}^* c_{i_3}$, $\rho_{i,q} = q_{i_3}^* c_{i_2} + q_{i_4}^* c_{i_4}$,
- $\lambda_{i,q} = q_{i_1} c_{i_1} + q_{i_2} c_{i_2}$, $\mu_{i,q} = q_{i_1} c_{i_3} + q_{i_2} c_{i_4}$, $\nu_{i,q} = q_{i_3} c_{i_1} + q_{i_4} c_{i_2}$, $\xi_{i,q} = q_{i_3} c_{i_3} + q_{i_4} c_{i_4}$,
- $\mathbf{c}_i = [c_{i_1}, c_{i_2}, c_{i_3}, c_{i_4}]^T$, $\mathbf{J}_{i,p}(\boldsymbol{\theta}_{i,p}) = \begin{bmatrix} p_{i_1} & p_{i_2} \\ p_{i_3} & p_{i_4} \end{bmatrix}$, $\mathbf{J}_{i,q}(\boldsymbol{\theta}_{i,q}) = \begin{bmatrix} q_{i_1} & q_{i_2} \\ q_{i_3} & q_{i_4} \end{bmatrix}$,
- $\boldsymbol{\theta}_{i,p} = [p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}]^T$ et $\boldsymbol{\theta}_{i,q} = [q_{i_1}, q_{i_2}, q_{i_3}, q_{i_4}]^T$.

L'algorithme de calibration robuste que nous proposons est résumé ci-dessous.

Algorithme: Calibration robuste basée sur un modèle Gaussien composé

entrée : D, M, B, $\mathbf{C}_i$, β_i, $\mathbf{y}$
sortie : $\hat{\boldsymbol{\theta}}$
initialisation: $\hat{\boldsymbol{\Omega}} \leftarrow \boldsymbol{\Omega}_{\text{init}}$, $\hat{\boldsymbol{\tau}} \leftarrow \boldsymbol{\tau}_{\text{init}}$, $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_{\text{init}}$
while *critère d'arrêt non atteint* **do**
  while *critère d'arrêt non atteint* **do**
1 Étape E : $\hat{\mathbf{w}}_i$ donné par (16), $i = 1, \ldots, D$
2 Étape M : $\hat{\boldsymbol{\theta}}_i$, $i = 1, \ldots, D$
    while *critère d'arrêt non atteint* **do**
3 $\hat{\boldsymbol{\theta}}_{i,p}$ donné par (18), $p = 1, \ldots, M$
4 $\hat{\boldsymbol{\Omega}}$ obtenu par (10) puis (11)
5 $\hat{\boldsymbol{\tau}}$ obtenu par (9)

Afin d'évaluer les performances de notre algorithme, nous souhaitons le comparer à d'autres approches de l'état de l'art : le cas classique Gaussien [14] et la calibration robuste basée sur une

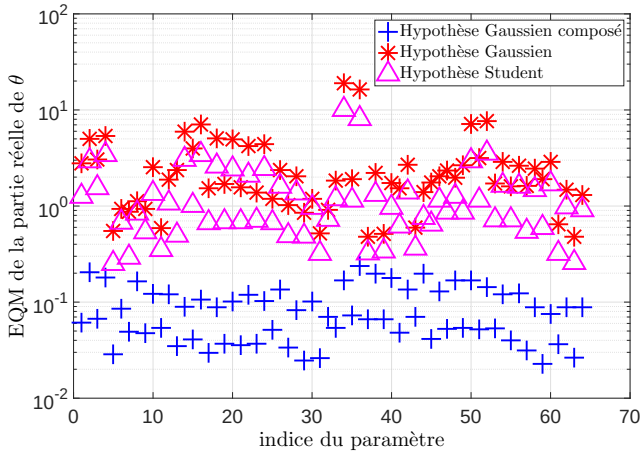


Figure 1: EQM de la partie réelle des 64 paramètres inconnus pour un RSB fixé, avec $D = 2$, $M = 8$ et $D' = 8$.

loi de Student [2], fournie par la littérature. Pour cela, nous nous plaçons dans un contexte réaliste avec la présence de D' sources de très faible intensité par rapport aux D sources de calibration connues. Sur la Figure 1, nous traçons l'Erreur Quadratique Moyenne (EQM) de la partie réelle de tous les paramètres inconnus $\boldsymbol{\theta}$, à Rapport Signal sur Bruit (RSB) fixé pour i) l'algorithme de calibration robuste proposé basé sur modèle Gaussien composé, ii) l'algorithme de calibration robuste traditionnel basé sur une loi de Student et iii) le cas Gaussien basé sur la résolution d'un moindres carrés ordinaire. Nous remarquons de meilleurs performances d'estimation dans notre cas pour l'ensemble des paramètres. En effet, contrairement à la littérature, aucune distribution spécifique n'a eu besoin d'être décrite et les entrées du vecteur bruit ne sont pas supposées indépendantes.

Matrices de Jones structurées

Une matrice de Jones peut se décomposer comme le produit de plusieurs matrices, chacune représentant un effet physique bien particulier. Dans le cas d'un régime de calibration appelé régime 3 [3], exposé en Figure 2, les antennes du réseau sont relativement proches les unes des autres et admettent chacune un très large champ de vision, comme c'est le cas à l'échelle d'une station du LOFAR ou du SKA [4]. Dans un tel contexte, nous considérons la décomposition suivante pour les matrices de Jones [25, 26, 31]

$$\mathbf{J}_{i,p}(\boldsymbol{\theta}_{i,p}) = \mathbf{G}_p(\mathbf{g}_p)\mathbf{H}_{i,p}\mathbf{Z}_i(\varphi_i)\mathbf{F}_i(\vartheta_i) \tag{19}$$

où $i \in \{1, \ldots, D\}$, $p \in \{1, \ldots, M\}$ et $\boldsymbol{\theta}_{i,p} = [\vartheta_i, \varphi_i, \mathbf{g}_p^T]^T$.

La seule matrice supposée connue ici est $\mathbf{H}_{i,p}$, qui tient compte du diagramme de rayonnement des antennes ainsi que du retard de propagation géométrique. Elle peut être supposée connue grâce à des simulations électromagnétiques effectuées au préalable et à l'information disponible fournie par la position de l'antenne et la direction de visée [5, 24, 32]. Du fait de la structure compacte du réseau d'antennes considéré, la même partie de l'ionosphère est observée et les effets environnementaux sont supposés identiques pour l'ensemble des antennes. Plus précisement, le passage dans l'ionosphère entraîne un déphasage qui s'écrit

$$\mathbf{Z}_i(\varphi_i) = \exp\Big(j\varphi_i\Big)\mathbf{I}_2 \tag{20}$$

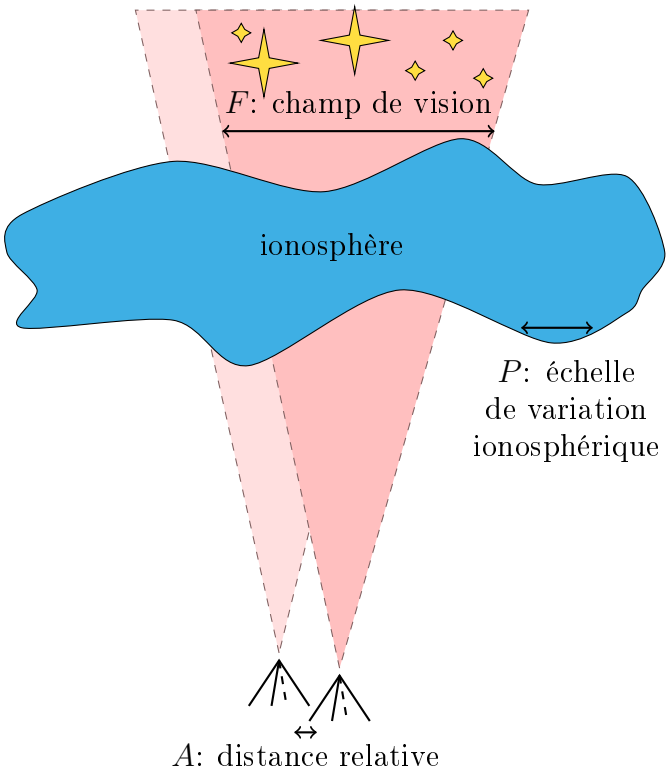


Figure 2: Régime de calibration 3, où $F \gg P$ et $P \gg A$.

ainsi qu'une rotation du plan de polarisation, appelée rotation de Faraday et donnée par

$$\mathbf{F}_i(\vartheta_i) = \begin{bmatrix} \cos(\vartheta_i) & -\sin(\vartheta_i) \\ \sin(\vartheta_i) & \cos(\vartheta_i) \end{bmatrix}. \tag{21}$$

Enfin, chaque antenne est décrite par un gain instrumental caractéristique de la chaîne de réception, tel que

$$\mathbf{G}_p(\mathbf{g}_p) = \mathrm{diag}\{\mathbf{g}_p\}. \tag{22}$$

Sur la Figure 2, nous exposons le principe du régime 3 : la même portion de l'ionosphère est considérée mais une multitude de sources peuvent être visibles dans le champ de vision. Notons que dans le cas où les antennes sont relativement éloignées les unes des autres et admettent toujours chacune un très large champ de vision, comme c'est le cas à l'échelle des plusieurs stations, il s'agit alors du régime 4 où nous pouvons introduire un déphasage ionosphérique $\varphi_{i,p}$ et un angle de rotation $\vartheta_{i,p}$ pour chaque source $i \in \{1, \ldots, D\}$ et chaque antenne $p \in \{1, \ldots, M\}$.

Après avoir estimé les entrées des matrices de Jones à l'aide de l'algorithme décrit dans le cas non structuré, il est possible de déduire les paramètres physiques qui interviennent dans (19). Pour cela, nous considérons une fonction coût de type moindres carrés et optimisons par rapport à chacun des paramètres inconnus de manière alternée. Ainsi, nous obtenons les estimations suivantes

1) Estimation de $\mathbf{g}_p$: Après calculs, chaque gain est donné par

$$[\hat{\mathbf{g}}_p]_k = \Big(\sum_{i=1}^{D} [\mathbf{W}_{i,p}^*]_{k,k} \Big)^{-1} \sum_{i=1}^{D} [\mathbf{X}_{i,p}^*]_{k,k} \tag{23}$$

où $\mathbf{X}_{i,p} = \mathbf{R}_{i,p}\hat{\mathbf{J}}_{i,p}^H$, $\mathbf{W}_{i,p} = \mathbf{R}_{i,p}\mathbf{R}_{i,p}^H$ et $\mathbf{R}_{i,p} = \mathbf{H}_{i,p}\mathbf{Z}_i\mathbf{F}_i$.

2) Estimation de φ_i : Après dérivation et égalisation à zéro, nous obtenons le résultat suivant

$$\exp\left\{2j\hat{\varphi}_i\right\} = \frac{\mathrm{Tr}\left\{\mathbf{M}_{i,p}\right\}}{\mathrm{Tr}\left\{\mathbf{M}_{i,p}^H\right\}} \tag{24}$$

où $\mathbf{M}_{i,p} = \sum_{p=1}^{M} \hat{\mathbf{J}}_{i,p}\mathbf{F}_i^H\mathbf{H}_{i,p}^H\mathbf{G}_p^H$ et $\hat{\varphi}_i$ peut directement en être déduit.

3) Estimation de ϑ_i : Une procédure numérique pour une dimension est nécessaire pour l'estimation des angles de Faraday. Cette dernière peut facilement être réalisée en un temps limité à l'aide d'algorithmes classiques de type Newton ou descente de gradient.

Pour le moment, nous nous sommes restreints au cas mono-fréquence. Or, en pratique, les systèmes d'observation actuels exploitent une large gamme de fréquences. Il nous semble donc essentiel de prendre en compte l'aspect multi-fréquentiel.

Méthode de calibration robuste multi-fréquentielle

Pour illustrer l'utilisation de notre algorithme dans un contexte multi-fréquentiel, nous choisissons ici de considérer le régime 4, où les stations d'antennes sont éloignées les unes des autres, et la décomposition en (19) peut s'écrire

$$\mathbf{J}_{i,p}^{[f]}(\boldsymbol{\theta}^{[f]}) = \mathbf{G}_p^{[f]}(\mathbf{g}_p^{[f]})\mathbf{H}_{i,p}^{[f]}\mathbf{Z}_{i,p}^{[f]}(\varphi_{i,p}^{[f]})\mathbf{F}_{i,p}^{[f]}(\vartheta_{i,p}^{[f]}) \tag{25}$$

où $\varphi_{i,p}^{[f]} \propto \frac{1}{f}$ et $\vartheta_{i,p}^{[f]} \propto \frac{1}{f^2}$ [25, 33]. Notons que l'adaptation de l'algorithme multi-fréquentiel au régime 3 peut directement être déduite à partir de l'étude proposée. Pour les gains électroniques $\mathbf{g}^{[f]}$, nous pouvons supposer une certaine régularité dans la variation fréquentielle à l'aide d'un modèle polynomial. Ainsi, nous introduisons

$$\mathbf{g}_p^{[f]} = \tilde{\mathbf{B}}^{[f]}\tilde{\mathbf{z}}_p \tag{26}$$

où $\tilde{\mathbf{B}}^{[f]} = \tilde{\mathbf{b}}^{[f]^T} \otimes \mathbf{I}_2$ et

$$[\tilde{\mathbf{b}}^{[f]}]_k = \left(\frac{f - f_0}{f_0}\right)^{k-1} \tag{27}$$

avec $k \in \{1, \ldots, \tilde{N}\}$ et une fréquence de référence f_0. L'ordre $\tilde{N}$ est choisi de manière arbitraire et les variables latentes $\tilde{\mathbf{z}}_p$ nécessitent d'être estimées.

En supposant des mesures indépendantes entre paires d'antennes mais aussi entre fréquences $f \in \mathcal{F} = \{f_1, \ldots, f_F\}$, la log-vraisemblance s'écrit de manière similaire à (8) en introduisant une somme supplémentaire sur toutes les fréquences. Aucune structure *a priori* n'est imposée sur la variation de $\tau_{pq}^{[f]}$ et $\boldsymbol{\Omega}^{[f]}$ en fonction de f. Pour résoudre un tel problème sous contraintes, nous considérons le Lagrangien suivant

$$L\left(\{\boldsymbol{\theta}^{[f]}\}_{f\in\mathcal{F}}, \mathbf{z}, \{\mathbf{x}^{[f]}\}_{f\in\mathcal{F}}\right) = \sum_{f\in\mathcal{F}} L^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right) \tag{28}$$

tel que $L^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right) = l^{[f]}\left(\boldsymbol{\theta}^{[f]}\right) + h^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right)$, où le terme d'attache aux données s'écrit

$$l^{[f]}(\boldsymbol{\theta}^{[f]}) = \sum_{pq} \frac{1}{\tau_{pq}^{[f]}} \mathbf{u}_{pq}^{[f]^H}(\boldsymbol{\theta}^{[f]})\boldsymbol{\Omega}^{[f]^{-1}}\mathbf{u}_{pq}^{[f]}(\boldsymbol{\theta}^{[f]}) \tag{29}$$

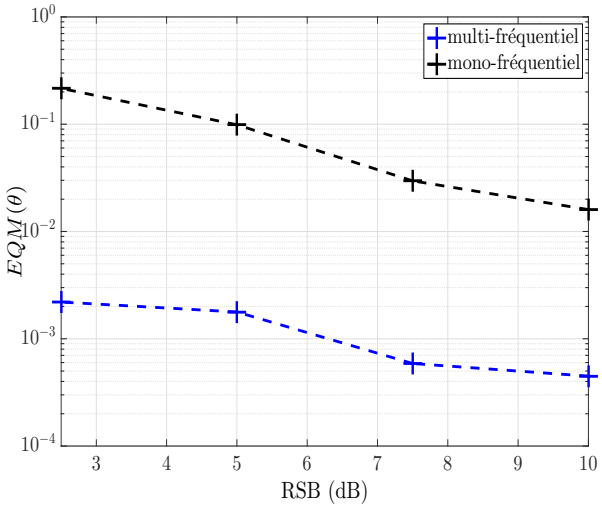


Figure 3: EQM d'un paramètre inconnu dans $\boldsymbol{\theta}$ en fonction du RSB.

et le terme de pénalisation comme suit

$$h^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right) = 2\Re\left\{\mathbf{x}^{[f]^H}\left(\boldsymbol{\theta}^{[f]} - \mathbf{B}^{[f]}\mathbf{z}\right)\right\} + \rho||\boldsymbol{\theta}^{[f]} - \mathbf{B}^{[f]}\mathbf{z}||_2^2. \tag{30}$$

Le vecteur $\mathbf{x}^{[f]}$ fait référence aux variables duales du problème pour une fréquence f, $\mathbf{z}$ inclut l'ensemble des variables cachées/latentes tandis que la matrice de couplage $\mathbf{B}^{[f]}$ tient compte des variations supposées connues des paramètres d'intérêt selon la fréquence. Pour résoudre un tel problème séparable en fréquence, nous utilisons une méthode dite distribuée qui permet de répartir la charge de calcul massive sur plusieurs agents [34, 35]. Ainsi, la calibration est réalisée localement au-niveau de chaque agent pour chaque fréquence tandis qu'un centre de fusion de l'information permet d'assurer la communication entre les différents agents, d'où une meilleure calibration possible puisque l'ensemble des fréquences est pris en compte et non pas une fréquence après l'autre de manière indépendante. Pour illustrer cela, nous montrons sur la Figure 3 l'évolution de l'EQM en fonction du RSB pour un paramètre représentatif des matrices de Jones dans deux cas spécifiques : i) calibration multi-fréquentielle avec une procédure distribuée et ii) calibration mono-fréquentielle où chaque fréquence est traitée de manière indépendante.

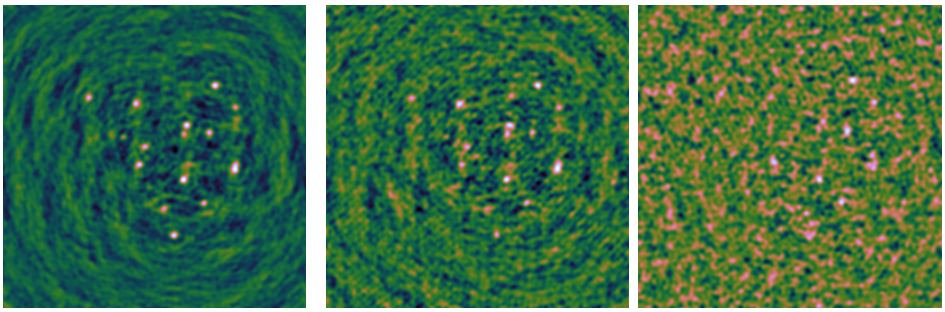
Une méthode appropriée pour résoudre un problème d'optimisation sous contraintes de manière distribuée est l'algorithme ADMM (*Alternating Direction Method of Multipliers*) [36] qui se compose de trois étapes successives, répétées itérativement,

- $\left(\hat{\boldsymbol{\theta}}^{[f]}\right)^{t+1} = \underset{\boldsymbol{\theta}^{[f]}}{\operatorname{argmin}}\, L^{[f]}\left(\boldsymbol{\theta}^{[f]}, (\hat{\mathbf{z}})^t, \left(\hat{\mathbf{x}}^{[f]}\right)^t\right)$ réalisée localement par chaque agent (31)

- $(\hat{\mathbf{z}})^{t+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{f\in\mathcal{F}} h^{[f]}\left(\left(\hat{\boldsymbol{\theta}}^{[f]}\right)^{t+1}, \mathbf{z}, \left(\hat{\mathbf{x}}^{[f]}\right)^t\right)$ réalisée globalement au centre de regroupement (32)

- $\left(\hat{\mathbf{x}}^{[f]}\right)^{t+1} = \left(\hat{\mathbf{x}}^{[f]}\right)^t + \rho\left(\left(\hat{\boldsymbol{\theta}}^{[f]}\right)^{t+1} - \mathbf{B}^{[f]}(\hat{\mathbf{z}})^{t+1}\right)$ réalisée localement par chaque agent (33)

Notons que l'étape (32) donne directement une expression analytique tandis que l'étape (31) nécessite une approche itérative en minimisant de manière alternée le Lagrangien $L^{[f]}$ par rapport



(a) (b) (c)

Figure 4: Images résiduelles obtenues avec (a) la réalité terrain (b) l'algorithme de calibration robuste proposé et (c) la calibration sous MeqTrees (mono-fréquence).

à chaque $\vartheta_{i,p}^{[f]}$, $\varphi_{i,p}^{[f]}$ et $\mathbf{g}_p^{[f]}$. Après calculs, des expressions analytiques peuvent être obtenues pour les gains électroniques ainsi que les déphasages ionosphériques. En ce qui concerne les angles de Faraday, une optimisation numérique est requise. Par souci de clarté, les expressions analytiques ainsi que le détail de l'optimisation alternée pour résoudre (31) sont donnés en section 4.3.2.

Simulations réalistes sous MeqTrees

Ce travail a aussi été l'occasion d'utiliser l'outil de simulation MeqTrees qui permet de générer des données interférométriques telles que celles mesurées par de réels instruments astronomiques [37]. Il permet notamment d'introduire un grand nombre de perturbations différentes dans le choix des matrices de Jones et propose également une méthode de calibration qui consiste à ajuster le modèle aux données. Généralement, nous nous intéressons aux images résiduelles : une fois la calibration réalisée en estimant les perturbations qui affectent les sources de calibration, ces dernières sont retirées des données pour finalement révéler la présence des sources de plus faible intensité, difficilement accessibles autrement. Sur la Figure 4, nous présentons un exemple d'images résiduelles obtenues avec $D' = 16$ sources de non-calibration ponctuelles et une calibration réalisée avec i) l'algorithme de calibration robuste que nous proposons et ii) le solutionneur disponible sous MeqTrees. Dans un premier temps, nous nous plaçons dans un contexte mono-fréquence.

En multi-fréquentiel, nous fournissons des résultats quantitatifs sur la Figure 5 en traçant l'évolution de l'amplitude restituée pour l'une des sources faibles de non-calibration en fonction de la fréquence.

Conclusion et perspectives

Ces travaux s'inscrivent dans le cadre de la calibration robuste pour les très grands radio-interféromètres. Les méthodes que nous avons proposées sont itératives, basées sur l'estimation au sens du maximum de vraisemblance et un modèle Gaussien composé pour le bruit, afin de tenir compte de la présence de sources inconnues de faible intensité et des différentes interférences. Afin d'accélérer les calculs, nous avons introduit différents algorithmes afin d'obtenir des expressions analytiques pour les paramètres d'intérêt. Une extension au cas multi-fréquentiel qui s'appuie sur l'ADMM a également été décrite. Des comparaisons avec des techniques de calibration classiques ont permis de révéler des résultats encourageants pour nos méthodes. Au cours d'un séjour à Nice, j'ai pu profiter de l'expérience de nos partenaires de l'ANR MAGELLAN, et plus

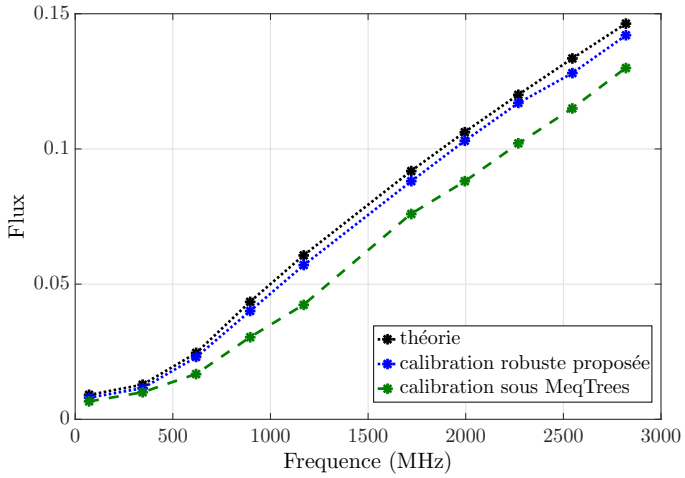


Figure 5: Restitution de l'amplitude en fonction de la fréquence.

particulièrement du Laboratoire J.L. Lagrange à Nice, pour travailler sur le logiciel de simulation MeqTrees et ainsi tester nos travaux d'estimation en calibration sur le processus d'imagerie de MeqTrees, le but étant de pouvoir révéler les sources faibles qui sont difficilement visibles.

Par la suite, il serait opportun d'adapter nos algorithmes à une échelle encore plus grande en ayant recours à des calculs très haute performance et à des architectures parallèles et ce, afin que l'implémentation sur données réelles puisse être pleinement accomplie. Enfin, des travaux méthodologiques sont également prévus, notamment sur des algorithmes d'estimation parcimonieux qui soient robustes ou encore des bornes de performances pour étudier la configuration optimale des antennes dans un réseau de capteurs.

Chapter 1

Introduction

1.1 Preamble

This thesis has been conducted as part of the research project ANR MAGELLAN whose aim is to develop algorithms for solving inverse problems of very large size. Such project gathers three different institutes: Observatoire de la Côte d'Azur (Lab. J.-L. Lagrange), Télécom ParisTech (LTCI) and ENS Cachan (SATIE). The global objective is to propose methodological tools for data processing in the context of advanced radio interferometers.

For sake of clarity and coherence, the work presented in this document does not account for the study made for separated arrays and structured noise covariance matrix **[C1]** as well as the derivation of Bayesian lower bounds in the Random Matrix Theory (RMT) framework **[C3]**.

1.2 Motivation of this work

Radio astronomy aims to study celestial objects at radio frequencies and this recent branch of astronomy has developed widely over the past century. Starting from the first detection of radio waves by Karl Jansky in 1928, it enabled to unveil new phenomena in space such as pulsars, quasars and radio galaxies, but also to supplement our knowledge by analyzing physical processes which are not detectable in optical astronomy. However, these discoveries require technological developments in the antenna receiving system which are referred to as radio antennas. They can be used individually or combined in a global array by exploiting interferometric techniques and aperture synthesis, thus, achieving high angular resolution. Modern radio interferometers include the existing Low Frequency Array (LOFAR) and the future Square Kilometre Array (SKA) which offer large collecting area, high level of sensitivity and unprecedented detailed and fast survey.

In order to reach the theoretical optimal performances of this new generation of radio interferometers, signal processing challenges must be overcome in terms of data correction and image synthesis. Indeed, estimation of all perturbation effects along the radio signal propagation path, namely the calibration process, is of critical importance in order to reconstruct images with no distortions. In this manuscript, we focus on the develoment of signal processing calibration algorithms, as it is a cornerstone of the imaging step in radio astronomy.

Array calibration strategies have already been proposed by the array processing community, based on the presence, or not, of cooperative sources, named calibration sources [38–40]. In radio astronomy, prior information is usually exploited thanks to tables describing accurately position and flux of the brightest sources. Still, antenna imperfections and disruptive environmental effects

hinder the observation of weak unknown background sources and thus, need to be corrected for. Most state-of-the-art calibration algorithms are iterative, based on a Gaussian noise assumption and operate in a mono-frequency scenario which reveal to be suboptimal and not always realistic in practice. Therefore, the aims are the following

- Investigate robustness aspects *w.r.t.* the presence of outliers or man-made radio frequency interferences, as the Gaussian noise model is usually considered in existing calibration algorithms. To the best of our knowledge, the only alternative robust algorithm using a non-Gaussian noise assumption is presented in [2, 41] and based on the Student's t noise modeling. Nevertheless, this modeling may not be the optimal way to incorporate the contribution of outliers in radio data as the distribution of such data is not known in practical scenario.

- Propose estimation algorithms with low computational complexity and efficient implementation.

- Study two major scenarios: calibration of interferometric systems with wide field-of-view and closely located (one station) or largely separated (multiple stations) radio antennas. These two different scales of observation refer, respectively, to calibration regimes 3 and 4, exposed in section 2.2.3 and are the most adapted for recent and next generation radio interferometers [3].

- Exploit the multi-frequency case where a wide frequency range is considered and efficient processing of data among all different frequencies is required.

- Characterize the estimation performances thanks to statistical studies and realistic reconstructed images in which weak unknown background sources can be revealed.

1.3 Achieved results

In this document, we propose robust calibration algorithms which are iterative, based on the Maximum Likelihood (ML) estimator and the Compound-Gaussian (CG) noise modeling. The CG distribution encompasses a wide range of distributions including the typical Gaussian, the Student's t, the Laplace and the Cauchy cases. Therefore, we choose to consider such flexible noise modeling to achieve robustness. In order to reduce the computational cost, efforts are being made to obtain closed-form expressions for all estimates when possible, by combining the use of the Expectation Maximization (EM) and Block Coordinate Descent (BCD) algorithms.

In radio astronomy, calibration amounts to estimate specific 2×2 matrices, called Jones matrices as described in section 2.3, which stand for all the introduced environmental and instrumental perturbation effects [5,18,26]. The first proposed robust algorithm in this document, namely the Relaxed Concentrated ML Estimator (RCMLE), operates in a mono-frequency scenario and is adapted for both non-structured and structured cases. This refers, respectively, to estimation of all entries of all Jones matrices and estimation of specific parameters in physical models (regime 3 as a first step). Statistical analysis with the Mean Square Error (MSE) and the Cramer-Rao Bound (CRB) highlight that the proposed scheme leads to more accurate estimation, thus, more robustness *w.r.t* the presence of weak unknown sources in the background and interferences. Let us note that the designed RCMLE is based on a relaxed version of the ML, *i.e.*, a deterministic assumption but its alternative, namely the Bayesian approach, is proposed as well in this document.

To extend our study to the multi-frequency case, we adopt a distributed approach with multiple local agents and one fusion center, unlike the classical strategy which uses a single centralized processor. We exploit the specific varition of parameters across frequency and conduct the study for the structured case. This leads to the proposed Multi-frequency Robust Calibration Algorithm (MRCA) designed for regime 4 with a possible straightforward adaptation to regime 3. Numerical simulations compare the estimation performances *w.r.t.* state-of-the art non-robust and/or mono-frequency cases (per-channel calibration).

Finally, to complete the analysis, we make use of the software MeqTrees for simulation of realistic radio data and visual indication of our improved performances in terms of calibration and image reconstruction [37].

1.4 Manuscript structure

The presented document relies mainly on the four following chapters

- Chapter 2 is dedicated to state-of-the-art calibration algorithms. We first present the context under consideration with a brief operating principle of interferometry and the new challenges at stake accompanying the development of advanced radio interferometers. The data model supported by Jones matrices is also introduced and most well-known calibration strategies are highlighted. A short description of the imaging procedure completes the chapter.

- Chapter 3 develops the proposed robust calibration algorithm based on th CG noise modeling. As a first step, the frequency dependence is not considered in this chapter.

- Chapter 4 adapts our methodological tools to the multi-frequency case. A distributed strategy is employed and robustness is still ensured.

- Lastly, chapter 5 introduces realistic data simulation with the software MeqTrees which helps us to visualize reconstructed residual images and reveal the presence of weak non-calibration sources, hidden in the noisy background.

1.5 Publications

The work described in this document has led to the following publications

Peer-reviewed international journals

[J1] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal and M. Pesavento, "Robust calibration of radio interferometers in non-Gaussian environment.", *IEEE* Transactions on Signal Processing, Volume : 65, Issue : 21, November 2017, pp. 5649-5660.
[J2] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer and P. Larzabal, "Robust distributed calibration of radio interferometers with direction dependent distortions.", accepted for publication by *Elsevier* Signal Processing, 2018.

Book chapter

[B1] Allen et al., "French SKA White Book - The French Community towards the Square Kilometre Array.", published by the SKA-France Coordination in collaboration with AS SKA-LOFAR, *Chapter 4 Technological developments*, 2017.

International conferences

[C1] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal and M. Pesavento, "Joint ML calibration and DOA estimation with separated arrays.", in Proc. *of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-16*, Shanghai, China.
[C2] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal and M. Pesavento, "Relaxed concentrated MLE for robust calibration of radio interferometers.", in Proc. of *the 24th European Signal Processing Conference, EUSIPCO-16*, Budapest, Hungary.
[C3] V. Ollier, R. Boyer, M. N. El Korso and P. Larzabal, "Bayesian lower bounds for dense or sparse (outlier) noise in the RMT framework.", in *the 9th Sensor Array and Multichannel Signal Processing Workshop, SAM-16,* Rio de Janeiro, Brazil.
[C4] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer and P. Larzabal, "Robust calibration of radio interferometers in multi-frequency scenario.", in Proc. *of IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP-18*, Calgary, Canada.
[C5] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer and P. Larzabal, "Bayesian calibration using different prior distributions: an iterative maximum a posteriori approach for radio interferometers.", in Proc. of *the 26th European Signal Processing Conference, EUSIPCO-18*, Rome, Italy.

National conferences

[C6] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal and M. Pesavento, "Algorithme de calibration robuste dans un contexte de radio interférométrie", in Proc. of *Colloque GRETSI*, 2017, Juan-Les-Pins, France.
[C7] V. Ollier, R. Boyer, M. N. El Korso and P. Larzabal, "Borne Bayésienne pour les systèmes large-échelle : cas d'un bruit à queue lourde", in Proc. of *Colloque GRETSI*, 2017, Juan-Les-Pins, France.

Technical report

[R1] V. Ollier, M. N. El Korso, R. Boyer and P. Larzabal, "Etat de l'art : méthodes de calibration en radioastronomie.", ANR MAGELLAN, Juin 2016, pp. 1-44.

Seminars

[P1] "Contributions aux méthodes de calibration robuste en radioastronomie", Journée des doctorants du laboratoire SATIE, 27 mai 2016, DIGITEO Shannon, Gif-sur-Yvette, France.
[P2] "A ML-based algorithm for calibration of radio interferometers" réunion ANR MAGELLAN, 21 juin 2016, Université de Nice, Nice, France.
[P3] "Relaxed concentrated MLE for robust calibration of radio interferometers", 3GC4 HI Fidelity Summer School, 31 octobre 2016, Port Alfred, South Africa.
[P4] "Robust Calibration of Radio Interferometers in Non-Gaussian Environment", Journée GDR ISIS, Traitement d'antenne : signaux non gaussiens, non circulaires, non stationnaires, 8 décembre 2016, Télécom ParisTech, Paris, France.
[P5] "La calibration robuste en radioastronomie", Journée département EEA, 26 janvier 2017, École Normale Supérieure Paris-Saclay, Cachan, France.
[P6] "Robust estimation of perturbation effects in radio astronomy", Séminaire au laboratoire CosmoStat, 2 février 2017, CEA Saclay, France.
[P7] "On robust calibration of radio interferometers: SIRP noise modeling and maximum likelihood estimator", Atelier SKA France, 27 février 2017, Université de Nice, Nice, France.

[P8] "Overview of robust calibration in radio astronomy", Journée des doctorants du laboratoire L2S, 15 juin 2017, CentraleSupélec, Gif-sur-Yvette, France.
[P9] "Calibration en radioastronomie par des méthodes de traitement statistique du signal", Journée des doctorants du laboratoire SATIE, 16 juin 2017, École Normale Supérieure Paris-Saclay, Cachan, France.
[P10] "Robust and distributed multi-frequency calibration method for radio interferometry, École d'été en traitement du signal et des images, 27 juin 2017, Peyresq, France.
[P11] "Méthodes de calibration robuste pour la radioastronomie", Journée des doctorants du laboratoire SATIE, 8 juin 2018, IFSTTAR, Versailles Satory, France.

Chapter 2

Problem setup and state-of-the-art

2.1 Context of study

The goal in radio astronomy is to measure radio emissions from the Milky Way galaxy with an antenna array, giving us more information about the formation of galaxies, stars, planets and the detection of pulsars and quasars. The first system to observe such wavelengths (9 kHz - 300 GHz) dates back to 1928 when Karl Jansky from Bell Telephone Laboratories mounted an antenna on a turntable. In the beginning, radio astronomy exploited single-antenna instruments with large dishes as the Arecibo observatory in Puerto Rico, completed in 1963 with 305 m in diameter and the Effelsberg 100 m radio telescope in Germany, unveiled in 1972 which is one of the largest fully steerable radio telescope.

With a single telescope to probe the sky, the corresponding angular resolution is given by

$$\theta = \frac{\lambda}{D} \tag{2.1}$$

where λ is the incident wavelength and D is the diameter of the instrument. Thus, a better resolution requires a larger diameter, which can involve cost and construction complexity issues.

To overcome these problems, the concept of interferometry was applied by considering an array with multiple antennas instead of one. The collected signals are combined so as to simulate a single large dish whose diameter would be equivalent to the largest distance between two antennas of the array, called the aperture. More specifically, the measured signals are correlated for each pair of antennas, spaced apart at a specific relative distance named baseline. The first radio observations based on interferometry occured in 1946 when Ryle and Vonberg used two dipoles. Nowadays, the most popular aperture synthesis arrays cover the Westerbork Synthesis Radio Telescope (WSRT), built in 1970 in the Netherlands and composed of 14 dishes of 25 m diameter each, or the Very Large Array (VLA) in the United States inaugurated in 1980 which consists of 27 radio telescopes with 25 m diameter as well. For such arrays, the angular resolution reads

$$\theta = \frac{\lambda}{A} \tag{2.2}$$

where A is the aperture. Therefore, finer resolution is attainable if the distance between antennas is increased. However, dishes, *i.e.*, parabolic antennas here, lack flexibility. In order to target a specific area in the sky, they need to be moved mechanically which may slow down the response time and bother observation of transient phenomena.

The new generation of radio interferometers is based on multiple flexible small antennas, omnidirectional, of low cost and with wide field-of-view. Contrary to mechanically based systems,

an observation area is selected thanks to phase shifts introduced electronically, *i.e.*, weighting by complex exponential terms. These antennas are called software telescopes: a beam is steered towards the direction of interest once signals captured by each individual antenna are correlated. Due to the huge number of available antennas in the array, a large collecting area is achieved and various different baselines are obtained, leading to better sensitivity, resolution, and a high quality image. An example of phased array telescope systems is the LOFAR [7,42,43] whose construction started in 2006 and was completed by the Netherlands Institute for Radio Astronomy (ASTRON). Regular observations have been conducted since 2012 and approximately 50 000 antennas are spread out in stations all over Europe, particularly in the Netherlands: more than 40 stations are built in the Netherlands, 5 in Germany, 1 in France, 1 in Sweden and 1 in Great-Britain. A station is composed of multiple individual antennas: 96 low band antennas (30-80 MHz) and 48 high band tiles (120-240 MHz) which are then combined by beamforming. For now, the LOFAR is the largest radio telescope ever built (based on interferometry) and the most sensitive in the radio frequencies (30-240 MHz). Furthermore, multiple observations can be conducted in parallel, up to 488 beams simultaneously.

The SKA [6, 8, 44] is the upcoming new generation of radio antennas, which will operate in a large observation bandwidth (up to tens of GHz) with a total collecting area of roughly one square kilometre. Expected to be the most sensitive and fastest astronomical instrument to probe the radio sky, its stations will be distributed in South Africa and Australia, composed of both traditional dishes and small dipoles. The construction is set to begin in 2018 and will be extended to its full size until 2030, in two consecutive phases. Precursor facilities of the SKA are already currently running: the Australian Square Kilometre Array Pathfinder (ASKAP), the Murchison Widefield Array (MWA) in Australia and the Karoo Array Telescope (MeerKAT) in South Africa.

2.1.1 Principles of interferometry

As mentioned before, interferometry consists in measuring correlations, called visibilities in radio astronomy, for two specific antennas, specifically distributed over an area at a given wavelength [26]. Due to Earth's rotation, baseline distances and orientations *w.r.t.* the sky field vary over time leading to more data points and high-resolution synthesis imaging. To give an idea, the longest baselines of the SKA will reach 3000 km at least, while the LOFAR attains 1000 km and the VLA, 36 km. By definition, a correlation V is given by the following expectation [19]

$$V(\mathbf{r}_1, \mathbf{r}_2) = \mathrm{E}\Big\{e(\mathbf{r}_1)e^*(\mathbf{r}_2)\Big\} \tag{2.3}$$

where, for sake of clarity, the electrif field e is assumed scalar and unpolarized, measured for two points in space, $\mathbf{r}_1$ and $\mathbf{r}_2$. For two locations, separated by a given propagation delay τ, one obtains [26]

$$V(\mathbf{r}_1, \mathbf{r}_2, \tau) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} e(\mathbf{r}_1, t)e^*(\mathbf{r}_2, t-\tau)dt. \tag{2.4}$$

In a suitable coordinate system, it is possible to link this spatial correlation with the source intensity. The relative distance $\mathbf{r}_1 - \mathbf{r}_2$ is usually expressed in units of wavelength, given by three coordinates (u, v, w), in which w is omitted if the coplanar approximation holds or the field-of-view is narrow. In the two-dimensional case, the visibility is given by the Fourier Transform of the source intensity, *i.e.*, [45]

$$V(u, v) = \int\int I(l, m)e^{-2\pi j(ul+vm)}dldm \tag{2.5}$$

where (l, m) refers to the angular coordinates of the source (right ascension (RA) and declination (DEC) for exemple), $I(l, m)$ is the intensity distribution at a given point and $V(u, v)$ is the measured correlation for a given baseline. In the ideal case, when the entire (u, v) plane is sampled by visibility measurements, the source intensity is given by the following Fourier inversion

$$I(l, m) = \int \int V(u, v) e^{2\pi j(ul+vm)} du dv. \tag{2.6}$$

The mathematical relationship in (2.5) is a particular expression of the fundamental Van Cittert-Zernike theorem [26, chp. 14], obtained by Van Cittert and Zernike [46], which is based on coherence theory and radiation emission by an incoherent electric field in the far-field. This theorem stems from the optical domain [47] and requires specific conditions: the emitting source is spatially incoherent, *i.e.*, the radiation emitted by one point of the source is statistically independent from any other point of the source (which is the case for most of the astronomical sources expect for pulsars and masers). In addition, the angular dimensions of the source must be typically inferior to one degree, the spectral width is quasi monochromatic and the source is in the far-field zone. It is worth mentioning that this relationship was initially obtained in the case of free-space between source and antennas, implying a uniform refractive index and an undistorted wavefront. Still, the generalization is achievable for an heterogeneous propagation medium and is referred to as Hopkins' formula.

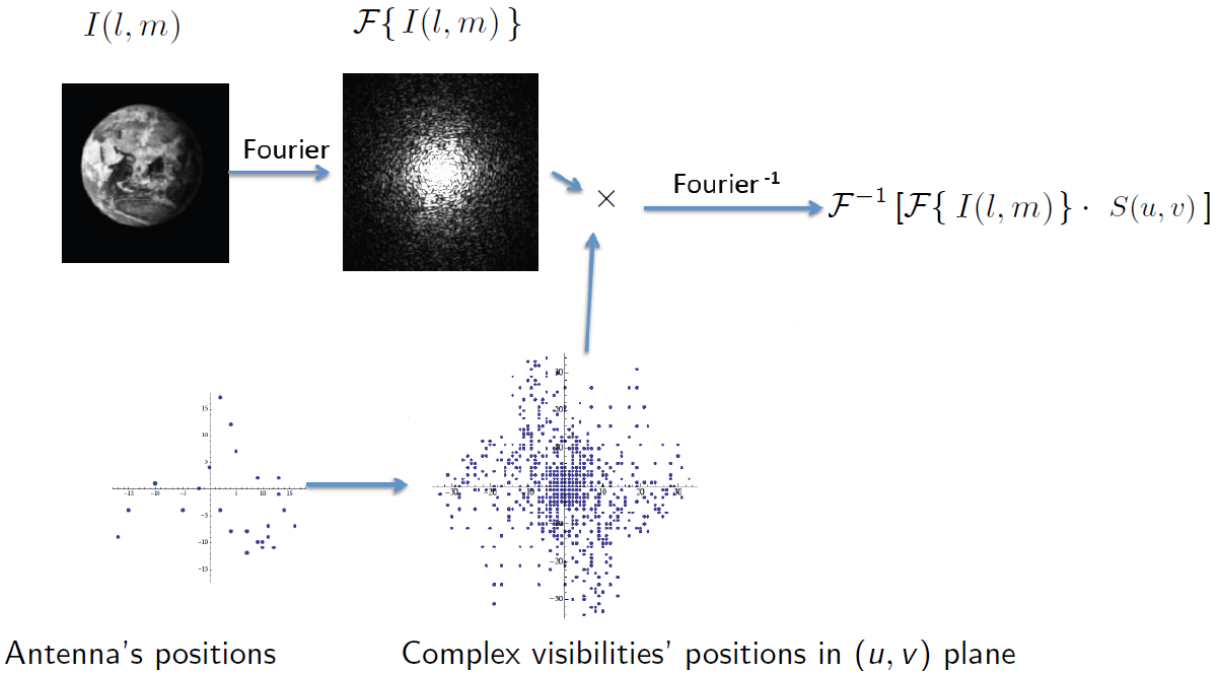


Figure 2.1: Data acquisition model with interferometry (after [1]).

The principle of data reconstruction with an interferometric system is exposed in Figure 2.1. The goal is to recover a faithful rendering of the original image, *i.e.*, $I(l, m)$. To this end, antennas are distributed over a specific layout in the array. Correlations are computed along each baseline vector between two antennas. Since the number of antennas is not infinite, only a limited set of baselines are achieved and correlations are not measured for any point of the space (u, v), leading to an inaccurate and distorted recovered image with missing information. Let us note $S(u, v)$ the sampling function representing actual positions for which correlations are computed, *i.e.*, a mask

which equals 0 when no measure has been conducted at this particular position of the plane or 1 elsewhere. If we multiply the two-dimensional Fourier Transform $\mathcal{F}\Big(I(l,m)\Big)$ with $S(u,v)$, the resulting data are the measured correlations. By Inverse Fourier Transform, we obtain an image, named the dirty image I_D, given by

$$I_D(l,m) = \int\int V(u,v)S(u,v)e^{2\pi j(ul+vm)}dudv. \tag{2.7}$$

As previously mentioned, synthesis mapping takes advantage of the Earth's rotation to generate a synthetic aperture. As the Earth rotates, the measured points in the (u,v) space follow elliptical tracks. The more data points we obtain, the more regular and dense is the sampling and the more accurate is the reconstructed image [48, p.24]. In Figure 2.2, we expose the uv-coverage, *i.e.*, the set of visibility points in the (u,v) domain, for different simulated instruments: the KAT-7 radio telescope composed of 7 dishes on the left hand side for a total measurement time of 12 hours, and the MeerKAT consisting of 64 dishes on the right hand side during 4 hours of observation. In both cases, the integration time per visibility point is 60 seconds and simulations were performed with the Common Astronomy Software Applications (CASA) package [49–51].

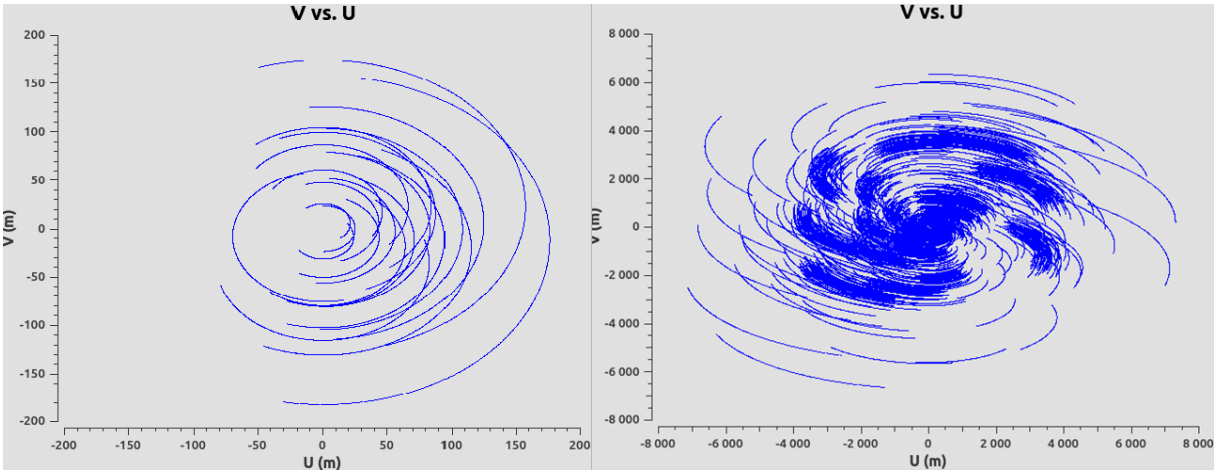


Figure 2.2: Different uv-coverage for different instruments: the KAT-7 radio telescope (left) and the MeerKAT instrument (right).

2.1.2 Challenges of the new generation of radio interferometers

Advanced phased arrays as the LOFAR and the SKA provide considerable benefits such as huge collecting area and large spatial coverage, fine sensitivity and flexibility, the possibility to observe in multiple directions simultaneously, a quick response time, ... However, great benefits necessarily produce some difficulties. Indeed, new challenges are at stake: a huge number of antennas with wide field-of-view, looking through an heterogeneous propagation medium, leads to significant direction dependent perturbation effects. Besides, the phased-array observational system has multi-frequency and multi-beam forming capabilities, polarization of waves needs to be considered and the theoretically achievable angular, temporal and spectral resolution is high. Antennas being ominidirectional and particularly sensitive, they can detect an important number of sources in their field-of-view. And in some cases, the source of interest reveals to be more complex than a simple point source, with constant intensity. Indeed, its structured diffuse

emission and its polarization state may change according to the frequency channel and along the signal propagation path.

With wide field-of-view and well-separated stations, the next generation array cannot be considered coplanar and spherical geometry is required. Indeed, all antennas are not located in a given plane at the same altitude. In this particular case, spatial coherency and brightness distribution are no longer linked thanks to the two-dimensional Fourier Transform in (2.5), since the coordinate w is essential and curvature of the Earth cannot be neglected. To solve these issues, faceting-based methods have been introduced as well as projection methods [49,50,52].

Last but not least, the new radio instruments are particularly sensitive and the amount of data to process is tremendous. We need to deal with large volume of data depending on the RA, the DEC and the frequency[1].

Therefore, to meet the theoretical optimal performances, many signal processing challenges need to be overcome, among which calibration, imaging and data processing [4, 9, 53]. In the following, we focus on the calibration step. Indeed, visibilities have assumed to be ideal so far, *i.e.*, uncorrupted. However, in practice, observed visibilities differ from the true visibilities as instrumentation is not perfect and physical disturbing phenomena are occuring. Thus, correcting these effects is essential in order to produce high dynamic range images with advanced observation systems.

2.2 Importance of calibration

2.2.1 Perturbation effects

When an incoming electromagnetic field reaches an array, it has been affected by various perturbation effects and physical distortions which involve propagation delay and magnitude loss. Here, we draw a distinction between, on the one hand, instrumental and electronic effects linked to the receiver chain and on the other hand, environmental effects due to propagation within the atmosphere.

• Instrumental effects: each antenna is defined by an unknown electronic gain and anisotropic beam pattern due to direction dependency and selectivity leakage. Composed of a main beam and sidelobes, the beam pattern accounts for sensitivity of antennas to radiation from any viewing direction. Electromagnetic simulations and modeling provide some a priori knowledge about it. However, because of sidelobes, an antenna is also sensitive in directions different from the one of interest. Thus, bright sources can hinder the observation of faint sources, which is particularly troublesome in the case of the LOFAR owing to its relatively large sidelobes. In addition, let us note that the beam pattern differs per station, *w.r.t.* time and frequency due to Earth's rotation. In the receiver chain, temperature may vary as well and other properties along the signal propagation path are changing through time until reaching the data processing center, thus affecting calibration parameters as gains.

• Environmental effects: the troposphere and the ionosphere are specific layers of the Earth's atmosphere. They entail time delays due to temporally and spatially varying diffraction and refraction phenomena, but also effects on the polarization of the signal as Faraday rotation [54] which is a rotation of the polarization plane of the wave while it goes through a charged medium as the atmosphere. This rotation is due to different propagation speeds between the two polarized components of the incoming wave. One prevailing effect is thus propagation within the

[1] Let us note that astronomical objects are identified by equatorial coordinates: the RA and the DEC, equivalent to celestial coordinates: latitude and longitude.

ionosphere, the ionized upper part, which leads to significant rapidly changing phase delays. We specify this point in the following remark.

Remark: The ionosphere is a specific layer of the atmosphere, which introduces time and phase delays, resulting in a shifted position of the source. Thus, the apparent position differs from the exact location. These visible consequences are due to non-uniform refractive index in the propagation medium and spatial, temporal fluctuations in the electron density. More specifically, the sun emits radiations which induce partial ionization in the atmosphere through injection of charged particles. This is called the solar wind. Therefore, free electrons emerge and slow the progression of electromagnetic waves. The movement of these free electrons is determined by the magnetic field, ionization, dissociation and ion recombination processes.
We can view the ionosphere as a dynamic, disruptive and heterogeneous environment which introduces phase shifts proportional to the wavelength, thus, inversely proportional to the frequency [21, 54]. This means that ionospheric perturbations are even more prevailing at low frequencies, especially for the LOFAR and the SKA which operate at low radio frequencies. Ionospheric disturbances are also temporally varying since at night, the number of electrons drops while it is higher during the day. Typically, at night and at the zenith, the number of electrons is given by 10^{16} electrons per square metre, which is equivalent to one unit of Total Electron Content (TEC). The TEC refers to the integral of electron density along the viewing direction and is directly related to the propagation time differences. Let us mention that correcting for artefacts due to the ionosphere remains a critical step in many scientific fields and not just in radio astronomy [55].

2.2.2 Aim of calibration

In radio astronomy, calibration is of the utmost importance before synthesis imaging and it directly deals with visibility measurements which may be affected by the perturbations, listed in section 2.2.1. In Figure 2.3, we show a simulated sky with undisturbed sources of varying intensity on the left hand side [2]. When the brightest sources suffer from perturbations, the resulting image is given on the right hand side. Thus, severe distortions and artefacts appear, making it difficult to reveal faint sources. These distortions are due to the simulated errors along the brightest sources but also to the incomplete Fourier sampling, leading to strong sidelobe patterns which overwhelm the image.

To have a chance of recovering the initial image, unknown perturbation effects need to be estimated and corrected. Afterwards, we can remove the contribution of the bright sources from the data, thanks to tables providing their position and intensity accurately [56, 57]. In the residual image, only the faintest sources remain visible which is of interest for the astronomers since these sources are the most difficult to access. Let us remark that most sources have an intensity lower than the noise level (except for particularly strong ones like the sun, Cygnus A (CygA) and Cassiopeia A (CasA) radio stars). Getting more knowledge about the sky model, the instrumentation and propagation conditions constitutes the ultimate goal of the calibration process.

2.2.3 Calibration regimes

Calibration of the ionosphere is addressed in different ways depending on the type of sensor array. Four different scenarios have been defined [3] and they are presented in Figure 2.4. They are distinguished by the aperture A, the size of the field-of-view F which is projected on the ionospheric layer and the scale P on which ionospheric perturbations are varying:

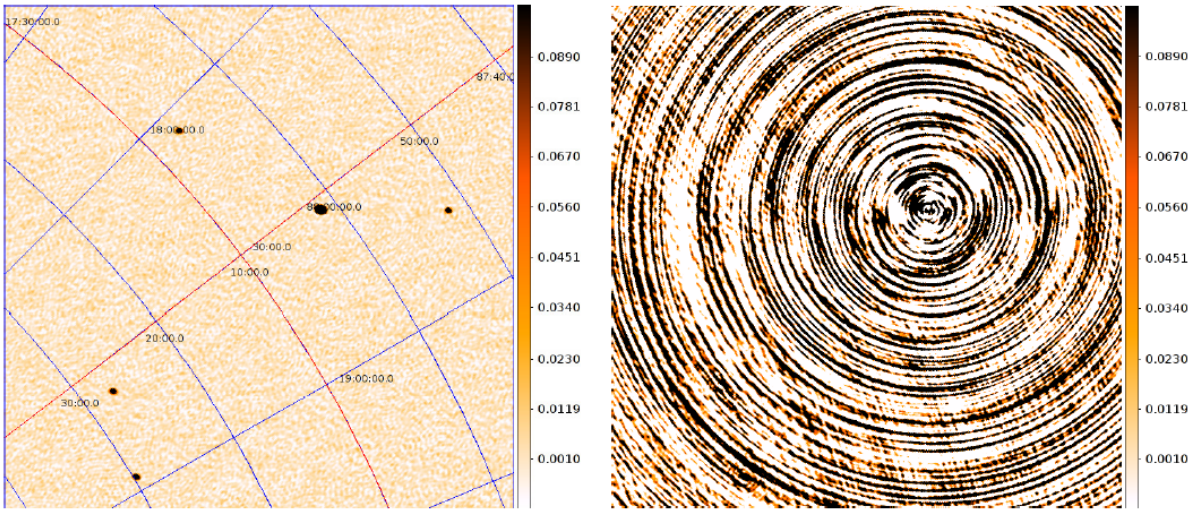


Figure 2.3: Comparison between an undistorted simulated sky (left) and a corrupted one (right) (after [2]).

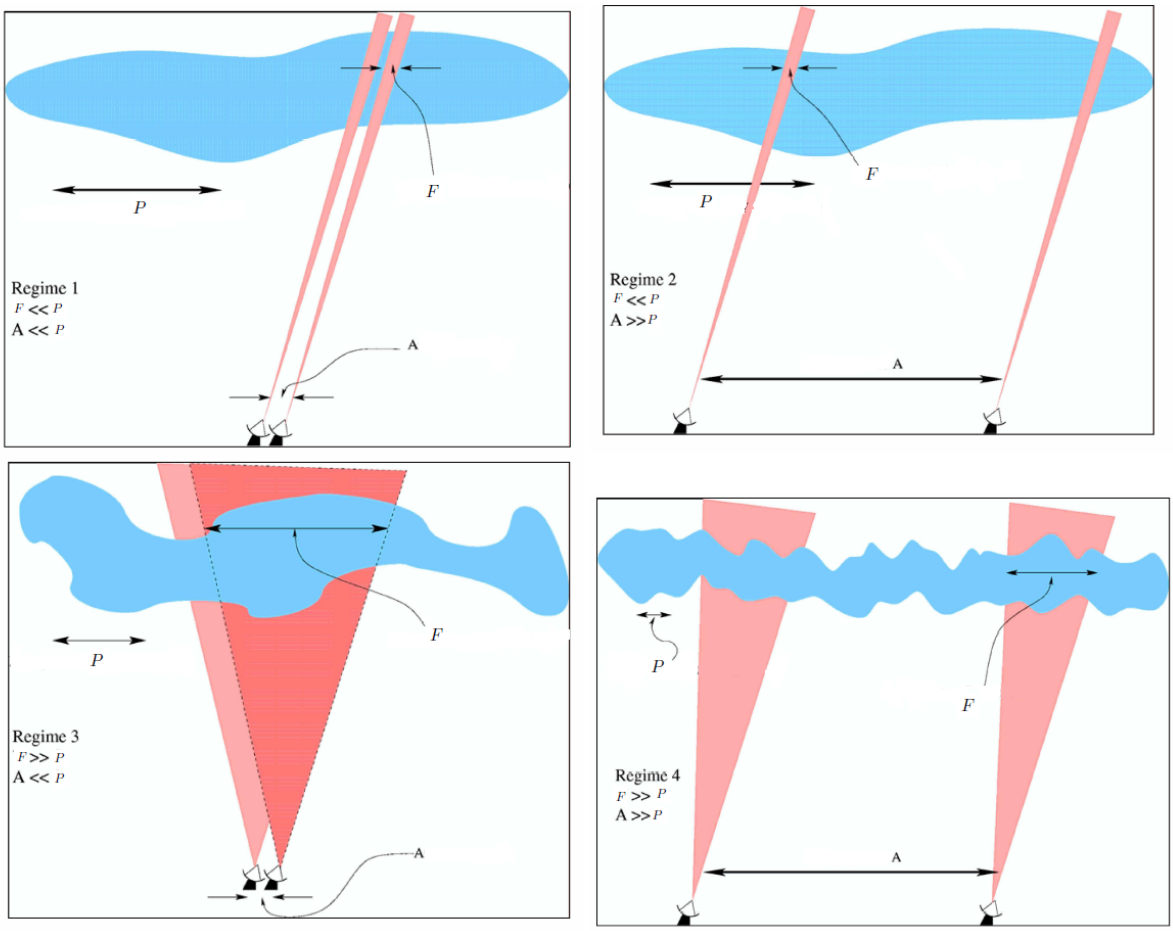


Figure 2.4: Ionospheric calibration regimes (after [3, 4]).

• In regime 1, the field-of-view F and the aperture A are very small in comparison with the ionospheric irregularity scale P. Consequently, each antenna sees the same part of the ionosphere so the same phase delay is introduced for each antenna in the array. The field-of-view being narrow, there is no direction dependency: the propagation conditions are the same for all paths within the field-of-view.

• In regime 2, the antennas are relatively separated from each other but their field-of-view remains narrow. When we target a specific direction, propagation conditions differ per antenna. Nevertheless, the same time delay is introduced for each source within the same field-of-view.

• Regime 3 refers to compact stations, with relatively closely spaced antennas, as in regime 1. By contrast, the field-of-view F is wide in comparison with the scale of variation P. Therefore, propagation conditions are direction dependent due to the wide field-of-view but antennas observe through the same part of the ionosphere, leading to similar perturbations. For instance, this amounts to consider a single station of the LOFAR or a subarray of the MWA.

• The most difficult case is regime 4 where stations are well separated from one another and their field-of-view are wide. Thus, the ionosphere induces per-antenna direction dependent perturbations.

In brief, ionospheric calibration requires to take into account variable properties, depending on the environment and the telescopes. The most adapted scenarios for the next-generation of radio interferometers are regimes 3 and 4 and different advanced approaches using array and signal processing techiques can be considered [58–60]. The conventional data model that is studied in radio astronomy is based on the so-called Jones matrices [19,24]. In the next section, we present this formalism for radio interferometer measurement equation [5, 18].

2.3 Data model

Let us consider a signal source with two components, each being associated to one specific polarization direction [61,62]. Unlike an unpolarized source which is defined by a single scalar, *i.e.*, its intensity, a polarized source is described by the distribution of the electric field. Thus, the wave nature of light needs to be considered. A priori information can be given about it thanks to survey observations. To describe the polarization state, we usually introduce the following 4 Stokes parameters [26]: (I, Q, U, V) with, respectively, an unpolarized, two linearly polarized and one circularly polarized components. This polarizarion of incident waves depends on position, frequency and can be severely affected by perturbations such as the Faraday rotation [54].

As mentioned before, the signal source is defined along two orthogonal polarization directions (horizontal and vertical or circularly right and left depending on the coordinate system), the third direction being the wave propagation direction. To retrieve both information, each antenna of the array is made up of two receptors (or dipoles), each being sensitive to one specific polarization direction, as shown in Figure 2.5. Thus, all the study is conducted in the plane orthogonal to the wave propagation direction. The incoming radio signal is decomposed as a 2×1 vector for the i-th source direction:

$$\mathbf{e}_i = [e_{i_x}, e_{i_y}]^T. \tag{2.8}$$

The incident electric field $\mathbf{e}_i$ and the measured voltage $\mathbf{v}_{i,p}(\boldsymbol{\theta})$ are related in the following way

$$\mathbf{v}_{i,p}(\boldsymbol{\theta}) = \mathbf{J}_{i,p}(\boldsymbol{\theta})\mathbf{e}_i \tag{2.9}$$

where the 2×2 Jones matrix, noted $\mathbf{J}_{i,p}(\boldsymbol{\theta})$, stands for all physical perturbations along the signal propagation path i-p, *i.e.*, from the i-th emitting source to the p-th receiving antenna,

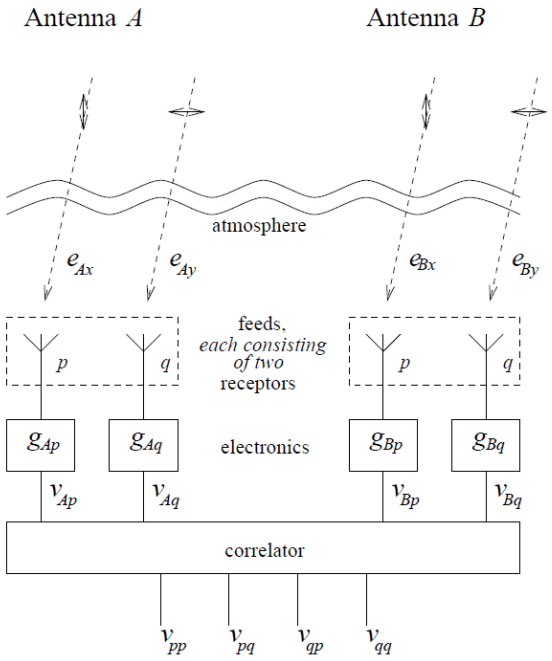


Figure 2.5: Interferometric system (after [5]).

and is parametrized by unknown vector $\boldsymbol{\theta}$. The relationship is given by a multiplication as propagation of the wave, and thus all undergone transformations, are assumed linear. We note $\mathbf{v}_{i,p}(\boldsymbol{\theta}) = [v_{i,p_x}(\boldsymbol{\theta}), v_{i,p_y}(\boldsymbol{\theta})]^T$ the generated voltage at the p-th dual polarized antenna and we can associate one particular Jones matrix $\mathbf{J}_{i,p}$ with each signal propagation path i-p.

The principle of an interferometer is to measure correlations between two specific antennas p and q. The resulting 2×2 ideal correlation matrix is given by

$$\mathbf{V}_{i,pq}(\boldsymbol{\theta}) = \mathrm{E}\Big\{\mathbf{v}_{i,p}(\boldsymbol{\theta})\mathbf{v}_{i,q}^H(\boldsymbol{\theta})\Big\} = \mathbf{J}_{i,p}(\boldsymbol{\theta})\mathbf{C}_i\mathbf{J}_{i,q}^H(\boldsymbol{\theta}) \tag{2.10}$$

where $\mathbf{C}_i = \mathrm{E}\Big\{\mathbf{e}_i\mathbf{e}_i^H\Big\} = \begin{bmatrix} I_i + Q_i & U_i + jV_i \\ U_i - jV_i & I_i - Q_i \end{bmatrix}$ stands for the intrinsic coherency or brightness of the i-th source, describing its polarization state. It is a function of previously introduced Stokes parameters, themselves functions of complex amplitudes of the electric field in a specific basis [18]. To take Jones matrices out of the expectation value in (2.10), they are assumed constant during the integration time interval.

Let us note that an alternative formulation of (2.10) exists, which makes use of the 4×4 Mueller matrices [50, p. 417], [63]. Instead of considering the matrix product between two voltage vectors, the outer product is used, leading to the following correlation vector

$$\mathbf{v}_{i,pq}(\boldsymbol{\theta}) = \Big(\mathbf{J}_{i,p}(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,q}^H(\boldsymbol{\theta})\Big)\Big(\mathbf{e}_i \otimes \mathbf{e}_i^H\Big) = \Big(\mathbf{J}_{i,p}(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,q}^H(\boldsymbol{\theta})\Big) \begin{bmatrix} I_i + Q_i \\ U_i + jV_i \\ U_i - jV_i \\ I_i - Q_i \end{bmatrix} \tag{2.11}$$

in which $\mathbf{J}_{i,p}(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,q}^H(\boldsymbol{\theta})$ describes the combine perturbation effects along the i-th signal path to antennas p and q. Such matrices are notably considered in [64] where the A-Projection deconvolution algorithm [65] is able to reconstruct an image, taking into account directional effects

as the frequency, time and polarization dependent beam pattern of antennas [10], and correcting for adverse effects. This particular 4×4 form is mathematically equivalent to the 2×2 one but is more adapted to depict imaging problems as it emphasizes the linear operations performed by the interferometer on the sky distribution [18]. In our study, we focus on the calibration procedure and the 2×2 formalism is thus favoured as it provides a more transparent, more intuitive and less complicated description of the underlying physics behind the signal propagation and it is also more suitable due to computational issues.

When the sky is composed of D multiple sources, assumed uncorrelated and far away, the emitted waves are considered planar when reaching the Earth. Thus, contributions of all sources are linearly added as follows

$$\mathbf{V}_{pq} = \sum_{i=1}^{D} \mathbf{J}_{i,p}(\boldsymbol{\theta}) \mathbf{C}_i \mathbf{J}_{i,q}^{H}(\boldsymbol{\theta}). \tag{2.12}$$

In a more realistic scenario, the contribution of calibration sources in the sky is not written as a discrete sum but rather like a diffuse radiation varying in direction. However, in this calibration study, we assume point sources as a first step. Besides, in practical scenario, we usually add an uncorrelated Gaussian noise $\mathbf{N}_{pq}$ in (2.12).

A given Jones matrix $\mathbf{J}_{i,p}$, for one specific propagation path $i-p$, can be decomposed into the multiplication of many individual terms, each being associated with one particular effect [24]:

$$\mathbf{J}_{i,p} = \mathbf{J}_{i,p,1} \mathbf{J}_{i,p,2} \cdots \mathbf{J}_{i,p,T}. \tag{2.13}$$

This is what we call a Jones chain [18], with T assumed different effects here. The arrangement is particularly important since it corresponds to the physical order in which effects appear along the path. Thus, the leftmost matrix $\mathbf{J}_{i,p,1}$ stands for the perturbation introduced belatedly, *i.e.*, close to the receiver, as the electronic gains, while the matrices on the right describe the first introduced perturbations close to the source emission, as the Faraday rotation for instance.

Various physical and geometrical effects can be described by Jones matrices as ionospheric phase shifts due to refraction phenomena but also rotation of the so-called parallactic angle between the reception system and the incident field or even change of coordinate system which requires transformation matrices (conversion into circular coordinates for instance can reveal to be more adapted in some cases and this choice depends on the specific design of telescopes). Let us also mention the possible interferences between antennas (since components of the polarized electromagnetic signal are not always well separated and one receptor may collect part of the radiation intended for the other receptor), the complex electronic gains of antennas, the filter bandwidth in the electronic chain and the errors due to the correlator [18, 24].

In the literature, Jones matrices have their own specific structures [24, 66]. Let us start with a general decomposition [19]

$$\mathbf{J}_{i,p} = \mathbf{G}_p \mathbf{B}_p \mathbf{D}_p \mathbf{E}_{i,p} \mathbf{P}_{i,p} \mathbf{K}_{i,p} \mathbf{T}_{i,p} \mathbf{Z}_{i,p} \mathbf{F}_{i,p} \tag{2.14}$$

in which $\mathbf{G}_p$ refers to the per-antenna diagonal electronic complex gain matrix and $\mathbf{B}_p$ is the electronic frequency bandpass. The direction independent matrix $\mathbf{D}_p$ stands for cross-leakage and interferences between receptors while direction dependent perturbations include the primary beam $\mathbf{E}_{i,p}$, a projection matrix $\mathbf{P}_{i,p}$ which takes into account the orientation of receptors *w.r.t.* the sky, the typical geometric pathlength difference $\mathbf{K}_{i,p}$, the tropospheric phase delay $\mathbf{T}_{i,p}$, the ionospheric phase delay $\mathbf{Z}_{i,p}$ and the ionospheric Faraday rotation $\mathbf{F}_{i,p}$. Particular effects represented by $\mathbf{G}_p$ and $\mathbf{E}_{i,p}$ are studied in [67], where geometry of dipoles is taken into account

within the LOFAR and a possible model of the beam pattern is needed. Besides, $\mathbf{E}_{i,p}$, $\mathbf{Z}_{i,p}$ and $\mathbf{F}_{i,p}$ are considered in [68]. In section 3.2.2, we will introduce the matrix $\mathbf{H}_{i,p}$ to account for all assumed known effects, including matrices $\mathbf{E}_{i,p}$ and $\mathbf{K}_{i,p}$. In the following, a detailed description of some of the above matrices with simple representations is given

• The matrix $\mathbf{K}_{i,p}$ expresses the propagation delay, *i.e.*, a phase shift due to geometrical path difference. Such phase shifts can be compensated by introducing additional electronic delays in the corresponding directions. This matrix which is proportional to identity, namely a scalar matrix, affects each component of the signal in the same way so there is independency *w.r.t.* polarization. Its expression is given by

$$\mathbf{K}_{i,p} = \exp\Big(- 2\pi j(u_p l_i + v_p m_i + w_p(n_i - 1))\Big)\mathbf{I}_2 \tag{2.15}$$

where the position of the p-th antenna is expressed in units of wavelength (u_p, v_p, w_p) and the direction of the i-th observed source is defined by three coordinates (l_i, m_i, n_i). Scalar matrices have the same representation in any coordinate system and can be switched to any place of the Jones chain.

• Another typical scalar Jones matrix expresses the phase shift $\varphi_{i,p}$ induced by the ionosphere. It is written as

$$\mathbf{Z}_{i,p} = \exp\Big(j\varphi_{i,p}\Big)\mathbf{I}_2. \tag{2.16}$$

• When no electronic cross-talk is assumed between sensors, the electronic gain matrix $\mathbf{G}$ is typically diagonal, *i.e.*,

$$\mathbf{G}_p = \mathrm{diag}\{\mathbf{g}_p\}, \tag{2.17}$$

meaning that each component is affected independently.

• A rotation matrix, as the Faraday rotation $\mathbf{F}_{i,p}$, is characterized by a given angle $\theta_{i,p}$ and written as

$$\mathbf{F}_{i,p} = \begin{bmatrix} \cos(\theta_{i,p}) & -\sin(\theta_{i,p}) \\ \sin(\theta_{i,p}) & \cos(\theta_{i,p}) \end{bmatrix}. \tag{2.18}$$

Let us note that the choice of coordinate system is important here since a rotation matrix becomes diagonal in a circular coordinate system [18].

Jones matrices have their own specific variations *w.r.t.* time, frequency and direction. Usually, effects on the left hand side of the Jones chain in (2.13) are the same for all sources within the field-of-view. Therefore, a matrix $\mathbf{J}_{i,p}$, for the specific propagation path $i - p$, can be decomposed into direction independent matrices $\mathbf{Q}_p$ and direction dependent effects $\mathbf{M}_{i,p}$, leading to

$$\mathbf{J}_{i,p} = \mathbf{Q}_p \mathbf{M}_{i,p} \mathbf{K}_{i,p} \tag{2.19}$$

where the scalar propagation delay $\mathbf{K}_{i,p}$ clearly appears since it is the one fundamental effect to remain in an interferometer even when observations are completely uncorrupted. Thus, we clearly distinguish direction independent effects linked to the antennas and the design of the instrumentation system, from direction dependent perturbations as the beam pattern, Faraday rotation, parallactic angle, propagation through the tropospheric and ionospheric layers [37,63], which are the most significant perturbations to calibrate in advanced telescopes nowadays [69,70].

As regards time and frequency, scale variations differ as well according to the Jones matrix. To give an example, the $\mathbf{B}_p$ bandwidth matrix mainly depends on frequency while the electronic gain $\mathbf{G}_p$ matrix is varying temporally essentially [37]. Let us mention that the ionospheric phase

delay is inversely proportional to frequency while Faraday rotation angle is inversely proportional to the square of frequency. These variations will be exploited in Chapter 4.

As it can be noticed from above, Jones matrices are a suitable representation to describe interferometric systems with different antennas and various propagation paths. Therefore, calibration amounts to estimate these specific matrices. Two options are possible in the estimation process: consider the combined effects in the global matrix $\mathbf{J}_{i,p}$ and directly estimate its entries (this is the non-structured case [14]) or estimate unknowns in a parametric model and introduce structured matrices as in the Jones chain (2.14) (the so-called structured case [70]). Considering a parametric model in the structured scenario can lead to misspecifications if the model does not correspond exactly to reality and estimation performance can easily deteriorate. However, such models introduce a few parameters to estimate and a smaller variance can be attained in the absence of misspecifications. On the other hand, the non-structured case can be more robust as no particular specification is made about the model but far more unknowns need to be estimated, leading to possibly worse statistical behavior. In this work, we deal with both cases and depending on the operating environment, the user can consider either scenario.

In the next section, we expose state-of-the-art calibration methods to correct all amplitude and phase errors due to the environment and the instrumentation, described by the previously introduced Jones matrices.

2.4 Calibration methods

General strategies have been developed in radio astronomy to reach accurate calibration and thus, high dynamic-range images. Two classical approaches are presented hereafter.

2.4.1 Calibration strategies: external and self-calibration

• External calibration: Conventional calibration consists in estimating unknown gains while observing an isolated and bright calibration source whose structure, *i.e.*, position and intensity, is given by tables. Source 3C286 is an example of external calibrator, used in the case of the VLA [71]. In practice, calibration *w.r.t.* a reference source needs to be repeated frequently in order to know the evolution of distortions through time. Besides, the calibration source must be close enough to the unknown target (the source of interest) *s.t.* incurred disturbances are similar, otherwise, interpolation is necessary [72]. Still, they must also be separated enough *s.t.* their contributions are clearly distinct in the data sets. Such method has some disadvantages: accurate a priori knowledge is required about properties of the calibration source and for antennas with wide field-of-view, information is only provided in the direction of the calibration source.

• Self-calibration: Also called self-cal, this approach is an enhancement of external calibration. Iteratively and alternatively, both the sky (parameters of interest) and different perturbations (nuisance parameters) are estimated, the two being mutually dependent. The self-cal procedure used to apply traditionally in regimes 1 and 2: there are no direction dependent effects but they may differ per antenna due to the instrumentation. This means that the field-of-view is narrow and the same part of the ionosphere is observed. However, extension to the direction dependent calibration problem has also been addressed [73].

The actual calibration and imaging strategy based on the self-cal and the LOFAR operation [6, 74] is represented in Figure 2.6: the sky model is upgraded in major and minor cycles while calibration parameters are optimized by a designed algorithm in the calibration cycle, which is the subject of this document. More specifically, thanks to a given sky model, we start from an

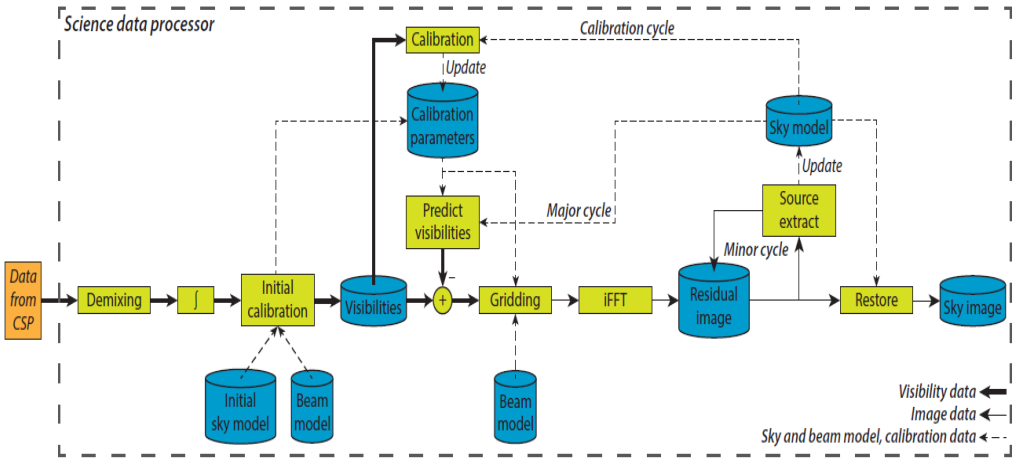


Figure 2.6: Science data processing in the LOFAR system (after [6]).

initial image which describes the intensity distribution, but some distortions and errors persist, as the reference model is not perfect. A Fourier Transform operator generates the corresponding correlations. Through the calibration cycle, perturbation parameters are estimated. Generally, they are calculated for the specific calibration source direction by minimizing the difference between measured correlations and those predicted by the model. But this point will be further specified. Once calibration is performed, a new reconstructed image leads to a new reference model. Therefore, progressively, as multiple back and forth are made between the correlations and the image data, the sky model is refined and estimation of calibration parameters gets more accurate.

To sum it up, the aim of self-cal was originally to estimate direction independent complex gains of antennas and noise parameters [75, 76]. But the problem has also been extended to multiple sources in the field-of-view [77, 78], where the assumed known source covariance matrix is equivalent to the source coherency matrix introduced in (2.10). Generalization to direction dependent perturbations, due to beam pattern or ionospheric effects, is also practicable [73, 79]. In this specific case, some assumptions are needed about the structure of the ionospheric layer or geometry of the array.

Finally, improved variants of the self-cal procedure are numerous: (demixed) Peeling [9,73,80] is adapted for direction dependent gains and calibration is performed *w.r.t.* the brightest source in the sky which is then subtracted from the data and the procedure is repeated for the next brightest source. Such technique has notably been deployed for the VLA and the MWA [21, 81]. We can also mention Pointing self-cal [82], which corrects for errors in pointing directions, leading otherwise to distortions in the recovered image. In addition, the Field-Based Calibration (FBC) is suitable for ionospheric calibration in regime 3 and estimates phase gradients [10, 83–85] using interpolation and Zernike polynomials. Indeed, in regime 3, antennas are relatively close to each other so the ionospheric variation within the field-of-view can be modeled as a linear gradient, using smooth polynomial model with low order. To generate such polynomials, Zernike or Karhunen-Loève transformations can be exploited. Another variant of Peeling is the Source Peeling and Atmospheric Modeling (SPAM) [86] which estimates ionospheric phase delays in regime 4. In this regime, antennas are well separated from one another and high order polynomials are more suitable to model the ionosphere.

In the next section, we give more details about array signal processing methods, which can be directly incorporated into the calibration cycle of Figure 2.6.

2.4.2 LS calibration

One of the most popular and statistically efficient method in signal processing is the ML estimator. However, it may suffer from heavy computational cost and it is not solvable in closed-form. An alternative option is to consider Least Squares (LS) based algorithms [12, 60] which are asymptotically efficient for large number of samples under Gaussian noise [87]. Besides, they benefit from reduced complexity *w.r.t.* the original ML and consist in minimizing the sum of squares of errors between data points and the predicted model. A possible nonlinear optimization technique to solve such problems is the gradient-based Levenberg-Marquardt (LM) algorithm [88], exposed briefly in the following.

2.4.2.1 LM algorithm

Using noisy observations from (2.12), its vectorization leads to

$$\text{vec}(\mathbf{V}_{pq}) = \mathbf{v}_{pq} = \sum_{i=1}^{D} \mathbf{J}_{i,q}^{*}(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,p}(\boldsymbol{\theta}) \text{vec}(\mathbf{C}_i) + \mathbf{n}_{pq} \tag{2.20}$$

with $\mathbf{n}_{pq} = \text{vec}(\mathbf{N}_{pq})$. Jones matrices are variable with time and frequency but as a first step, we are considering a sufficiently fine interval so that variations are neglected. Time t and frequency f dependence are thus not specified for now.

Considering only cross-correlations for all antenna pairs [14, 89], one obtains the following $4B \times 1$ vector, with $B = M(M-1)/2$ the total number of baselines,

$$\mathbf{y} = [\mathbf{v}_{12}, \mathbf{v}_{13}, \ldots, \mathbf{v}_{(M-1)M}]^T = \sum_{i=1}^{D} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n} \tag{2.21}$$

where $\mathbf{s}_i$ incorporates the contribution of known coherency of the i-th source and the Jones matrices for all baselines, *i.e.*, $\mathbf{s}_i(\boldsymbol{\theta}) = \left[\mathbf{s}_{i,12}^T(\boldsymbol{\theta}), \mathbf{s}_{i,13}^T(\boldsymbol{\theta}), \ldots, \mathbf{s}_{i,(M-1)M}^T(\boldsymbol{\theta})\right]^T$ and $\mathbf{s}_{i,pq}(\boldsymbol{\theta}) = \mathbf{J}_{i,q}^{*}(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,p}(\boldsymbol{\theta}) \text{vec}(\mathbf{C}_i)$.

The corresponding expression of the LS cost function is given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \, ||\mathbf{y} - \mathbf{s}(\boldsymbol{\theta})||_2^2 \tag{2.22}$$

in which $\mathbf{s}(\boldsymbol{\theta}) = \sum_{i=1}^{D} \mathbf{s}_i(\boldsymbol{\theta})$. If we note $\zeta(\boldsymbol{\theta}) = ||\mathbf{y} - \sum_{i=1}^{D} \mathbf{s}_i(\boldsymbol{\theta})||_2^2$ the considered cost function, the $(t+1)$-th iteration of the LM algorithm reads

$$\boldsymbol{\theta}^{t+1} = \boldsymbol{\theta}^t - \left(\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \zeta(\boldsymbol{\theta}) + \lambda \text{diag}\{\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \zeta(\boldsymbol{\theta})\}\right)^{-1} \nabla_{\boldsymbol{\theta}} \zeta(\boldsymbol{\theta})|_{\boldsymbol{\theta}^t}. \tag{2.23}$$

in which $\nabla_{\boldsymbol{\theta}}$ is the gradient with respect to $\boldsymbol{\theta}$ and $\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T$ is the Hessian matrix. Numerical implementation of the LM is exposed in [88, 90]. Let us note that a possible adaptation of the LM considers an identity matrix instead of the diagonal of the Hessian matrix, noted $\text{diag}\{\nabla_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}}^T \zeta(\boldsymbol{\theta})\}$ here. Any adapted positive-definite diagonal matrix can be used.

The LM algorithm [91, 92] is based on a damped Gauss-Newton method, *i.e.*, a combination between two minimization techniques: the gradient descent and the Gauss-Newton, depending

on how far the estimated parameters are from the optimal values. Choice of the damping parameter λ and initializations impact convergence and reaching the global optimum is not always possible. Due to computation of a Jacobian and inversion of large-dimensional matrices in (2.23), convergence can be extremely slow. The computational cost gets even more significant if the number of parameters to estimate is huge, as it is the case for very large radio telescopes. Speeding up the procedure is possible using Graphics Processing Unit (GPU) [93] or simplifying the problem with variants of the LS as EM or the Space-Alternating Generalized Expectation-Maximization (SAGE) algorithms, presented in section 2.4.3.

In the following section, we aim to present the (Weighted) Alternating Least-Squares ((W)ALS) algorithm which is asymptotically unbiased and efficient. It leads to closed-form expressions for some of the estimates and fulfills the convergence properties.

2.4.2.2 (W)ALS algorithm

In this subsection, we present array signal processing methods, based on the so-called covariance model. Such model is slightly different from the Jones model in (2.10) and was initially used to design the ALS and its variants. Nevertheless, the same approach can be applied to the Jones-based model.

Let us consider M antennas in an array with impinging D signal sources. Each antenna p with $p \in \{1, \dots, M\}$ is described by its position $\mathbf{l}_p = [x_p, y_p, z_p]^T$ which is known. Likewise, each source i with $i \in \{1, \dots, D\}$ is defined by its specific spatial location $\mathbf{d}_i = [l_i, m_i, n_i]^T$ *s.t.* $n_i = \sqrt{1 - l_i^2 - m_i^2}$ [60]. The $M \times D$ steering matrix is thus given by

$$\mathbf{A} = \exp\Big(- j\frac{2\pi}{\lambda}\mathbf{LD}\Big) \tag{2.24}$$

in which λ is the wavelength, $\mathbf{L} = [\mathbf{l}_1, \dots, \mathbf{l}_M]^T$ and $\mathbf{D} = [\mathbf{d}_1, \dots, \mathbf{d}_D]$. If the narrowband assumption holds true, geometric time delays account for phase shifts [94]. At each time instant, the signals measured by all antennas are collected in

$$\mathbf{r}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{2.25}$$

where $\mathbf{s}$ and $\mathbf{n}$ refer to the source and noise contribution, respectively. The noise is usually assumed to follow a zero-mean white Gaussian distribution. From a signal processing perspective, signal sources can be modeled in two different ways: either they are assumed deterministic or stochastic, *i.e.*, random and following a given a priori distribution.

In our specific application, interferometers compute correlations so we consider covariance matching estimation techniques. In the ideal case, when no perturbation occurs, except for the noise, the resulting covariance matrix reads [53]

$$\mathbf{R} = \mathrm{E}\Big\{\mathbf{r}(t)\mathbf{r}^H(t)\Big\} = \mathbf{A}\mathbf{S}\mathbf{A}^H + \mathbf{N} \tag{2.26}$$

where $\mathbf{S}$ is the source covariance matrix (known as the coherency/brightness matrix in the Jones-based model) with source powers along the diagonal and $\mathbf{N}$ is the noise covariance matrix which is diagonal in classical cases, with variable or not noise powers along the diagonal. Thus, the set of observations consists in covariance matrices which are functions of the source structure, the environmental and instrumental errors, and the receiver noise. The ALS approach is a suitable technique for solving LS covariance model fitting as follows

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, ||\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta})||_F^2 \tag{2.27}$$

where $\hat{\mathbf{R}}$ is the sample covariance matrix, *i.e.*, an estimate of the data covariance matrix, given by $T^{-1}\mathbf{Y}\mathbf{Y}^H$ with $\mathbf{Y} = [\mathbf{r}(t_1), \ldots, \mathbf{r}(t_T)]$, T the number of samples and $\boldsymbol{\theta}$ refers to all the unknown parameters in the model. The ALS consists in minimizing alternatingly the given cost function *w.r.t.* each unknown parameter, while fixing the others at their previous estimates, starting from a refined enough initialization. The choice of initializations is important in the ALS as convergence to the global optimum is not ensured. However, decrease of the cost function at each iteration step can be guaranteed. As for the WALS, it consists in using an appropriate weighting term $\boldsymbol{\Omega}$ *s.t.* the cost function becomes: $||\boldsymbol{\Omega}^H\Big(\hat{\mathbf{R}} - \mathbf{R}(\boldsymbol{\theta})\Big)\boldsymbol{\Omega}||_F^2$. Better statistical performance can be achieved with the WALS, at the expense of slight increase in the computational cost.

Depending on the calibration regime under study, the data model in (2.26) is different as well as the unknown parameters to estimate. Therefore, in what follows, we adapt the data model in (2.26) to the calibration regimes introduced in section 2.2.3

• Regime 1: In the general case, calibration requires to estimate direction independent electronic gains $\mathbf{G} = \text{diag}\{\mathbf{g}\}$ *s.t.*

$$\mathbf{R} = \mathbf{G}\mathbf{A}\mathbf{S}\mathbf{A}^H\mathbf{G}^H + \mathbf{N}. \tag{2.28}$$

These gains are the same for all sources in the field-of-view which is narrow. Consequently, in regime 1, we only need to observe a calibration source whose intensity and position are given. If we are pointing towards this specific source, the matrix $\mathbf{A}$ reduces to a vector $\mathbf{a}$ and $\mathbf{S}$ to the scalar source power σ_s^2 which are both known from tables, leading to the simplified problem

$$\mathbf{R} = \tilde{\mathbf{g}}\tilde{\mathbf{g}}^H + \mathbf{N} \tag{2.29}$$

where $\tilde{\mathbf{g}} = \mathbf{G}\mathbf{m}$ with $\mathbf{m} = \mathbf{a}\sigma_s$, known as a rank-1 factor analysis model from which $\mathbf{g}$ and $\mathbf{N}$ can be solved [13].

• Regime 2: The ionospheric perturbations are still the same for the entire field-of-view due to its narrowness but each individual antenna is assigned a unique perturbation term, leading to a similar calibration procedure as in Regime 1.

• Regime 3: Direction dependent effects and beam patterns of antennas need to be taken into account. To this end, we introduce direction dependent gains $\boldsymbol{\gamma}$ *s.t.*

$$\mathbf{R} = \mathbf{G}\mathbf{A}\boldsymbol{\Gamma}\mathbf{S}\boldsymbol{\Gamma}^H\mathbf{A}^H\mathbf{G}^H + \mathbf{N} \tag{2.30}$$

where $\boldsymbol{\Gamma} = \text{diag}\{\boldsymbol{\gamma}\}$. For the calibration sources, $\mathbf{S}$ is known and we usually consider $\boldsymbol{\Sigma} = \boldsymbol{\Gamma}\mathbf{S}\boldsymbol{\Gamma}^H$ as a global unknown diagonal matrix [78].

• Regime 4: A gain needs to be estimated for each source and each antenna, since many unknown sources are visible in the field-of-view. The data measurements are given by [73]

$$\mathbf{R} = (\mathbf{G} \odot \mathbf{A})\mathbf{S}(\mathbf{G} \odot \mathbf{A})^H + \mathbf{N} \tag{2.31}$$

where $\mathbf{G}$ is no longer diagonal, but full and considers all possible perturbation effects whether they are direction dependent or not. Such calibration problem requires further assumptions in order to be identifiable.

Advanced interferometric arrays are described by regimes 3 and 4 in which directional effects are prevailing. Therefore, we will focus much of our attention on these two regimes:

• In regime 3, we need to estimate direction independent instrumental gains $\mathbf{G}$, the source powers $\mathbf{\Sigma} = \text{diag}\{\boldsymbol{\sigma}\}$ and the noise powers $\mathbf{N} = \text{diag}\{\boldsymbol{\sigma}_n\}$ as follows [12,95]

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\text{argmin}} \, ||\hat{\mathbf{R}} - \mathbf{G}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^H\mathbf{G}^H - \mathbf{N}||_F^2 \tag{2.32}$$

$$\hat{\boldsymbol{\sigma}} = \underset{\boldsymbol{\sigma}}{\text{argmin}} \, ||\hat{\mathbf{R}} - \mathbf{G}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^H\mathbf{G}^H - \mathbf{N}||_F^2 \tag{2.33}$$

$$\hat{\boldsymbol{\sigma}_n} = \underset{\boldsymbol{\sigma}_n}{\text{argmin}} \, ||\hat{\mathbf{R}} - \mathbf{G}\mathbf{A}\mathbf{\Sigma}\mathbf{A}^H\mathbf{G}^H - \mathbf{N}||_F^2. \tag{2.34}$$

Equation (2.32) requires a rank-1 approximation in order to get an estimate while (2.33) and (2.34) lead to closed-form expressions. In fact, the gain estimation step can be addressed in various ways: iteratively with a Gauss-Newton method (GNLS) [96], logarithmically with the Logarithmic Least Squares (LOGLS) [13] or even using Column Ratios (COLR) [12,95].

Powers and positions of the calibration sources, as well as antenna locations are typically assumed known. Due to the possible large number of unknowns to estimate, constraints can be imposed on a reference source with fixed power, phase and position in order to avoid any identifiability issues. However, due to ionospheric phase shifts, exact positions of calibration sources may still be known inaccurately. To address this problem, a Direction Of Arrival (DOA) estimation step can be added in the iterative procedure. These inaccurate location parameters appear in $\mathbf{A}$ whose parametric model is available, and are estimated with MUSIC or Weighted Subspace Fitting (WSF) algorithms [97], which are asymptotically statistically efficient and make use of signal and noise subspaces [98]. Incorporating the DOA estimation in the iterative procedure leads to an extension of ALS, named xALS [12,95].

• In regime 4, the number of unknowns is tremendous. To be able to solve the problem and remove identifiability issues, physical constraints must be added. These constraints can be related to the structured array into multiple compact subarrays (regime 3) or the specific variation of perturbations *w.r.t.* time and frequency, with a difference between instrumental gains and ionospheric effects. Imposing constraints enables to reduce the number of unknown parameters to estimate so that the Fisher Information Matrix (FIM), whose inverse leads to the CRB, is invertable with no singularities.

If we exploit time t and frequency f, the model (2.31) can be written as [73]

$$\mathbf{R}_t^{[f]} = (\mathbf{G}_t^{[f]} \odot \mathbf{A}_t^{[f]})\mathbf{S}_t^{[f]}(\mathbf{G}_t^{[f]} \odot \mathbf{A}_t^{[f]})^H + \mathbf{N}_t^{[f]} \tag{2.35}$$

and the calibration problem is considered for each time interval and frequency band, as follows

$$\hat{\boldsymbol{\theta}}_t^{[f]} = \underset{\boldsymbol{\theta}_t^{[f]}}{\text{argmin}} \, ||\hat{\mathbf{R}}_t^{[f]} - \mathbf{R}_t^{[f]}(\boldsymbol{\theta}_t^{[f]})||_F^2 \;\; \text{s.t.} \;\; C\big(\{\boldsymbol{\theta}_t^{[f]}\}_{t,f}\big) = 0 \tag{2.36}$$

where C stands for the considered constraints *w.r.t.* time and frequency.

In the following, we describe the Statistically EFficient and Fast Calibration (StEFCal) method, adapted for antenna-based gains calibration [61,62], which provides computational advantage over the previously introduced algorithms derived from (weighted) LS cost function.

2.4.2.3 StEFCal algorithm

Introduced in [61], this Alternating Direction Implicit (ADI) approach proposes to reduce the numerical complexity of typical LM solver which is restraining for implementation in modern radio interferometers. Its convergence to optimal solution is guaranteed and it reveals to be more efficient than traditional antenna-based gain methods.

In this method, the gain of each receive path is estimated, assuming that gains of all other antennas are already calibrated. The iterative procedure is then repeated successively for each antenna. Such methods have already been considered in [62, 81, 99]. To improve convergence, robustness and reduce the influence of noise on the estimates, numerical simulations have shown that averaging even and odd iterations accelerates the convergence. Powers and positions of sources being known, unknown parameters to estimate reduce to the complex gains describing the elements of the array and receiver path noise powers, leading to the following calibration problem

$$\{\hat{\mathbf{g}}, \hat{\boldsymbol{\sigma}}_n\} = \underset{\mathbf{g},\boldsymbol{\sigma}_n}{\operatorname{argmin}} ||\boldsymbol{\Omega}^H(\hat{\mathbf{R}} - \mathbf{GASA}^H\mathbf{G}^H - \mathbf{N})\boldsymbol{\Omega}||_F^2 \tag{2.37}$$

where $\boldsymbol{\Omega}$ is a possible weighting term.

Autocorrelations being essentially dominated by noise, estimation of gains is performed using cross-correlations only. Thus, to simplify the problem during the gain calibration step, the noise $\mathbf{N}$ is ignored and the diagonals of $\hat{\mathbf{R}}$ and $\mathbf{M}$ are set to 0. We note $\mathbf{M}$ the available model of the observed field, which is approximately equal to $\mathbf{ASA}^H$ and only includes the undistorted brightest calibration sources. The estimation problem is then described by

$$\hat{\mathbf{g}} = \underset{\mathbf{g}}{\operatorname{argmin}} ||\hat{\mathbf{R}} - \mathbf{GMG}^H||_F^2. \tag{2.38}$$

The underlying approach consists in estimating $\mathbf{g}^H$ for fixed $\mathbf{g}$, and vice versa. Considering decoupled complex gains, the objective function in (2.38) can be written as

$$||\hat{\mathbf{R}} - \mathbf{GMG}^H||_F^2 = \sum_{p=1}^{M} ||[\hat{\mathbf{R}}]_{:,p} - [\mathbf{GMG}^H]_{:,p}||_F^2. \tag{2.39}$$

Therefore, the problem is divided into independent linear LS problems and each sensor gain is updated successively for $p \in \{1, \ldots, M\}$ during one given iteration.

Statistical estimation performance is similar whether we use directly the LM as an optimization technique or the StEFCal procedure. Still, this last method results in lower complexity since $\mathcal{O}(M^2)$ is required instead of $\mathcal{O}(M^3)$. Variants of StEFCal are exposed in [61]. For instance, to reduce the effects of outliers, a ponderation term can be incorporated leading to the Iterative Reweighted Least Squares (IRLS).

The StEFCal method can also be directly used as the gain estimation step of the ALS procedure in (2.32) so as to improve the implementation, the computational cost and the speed of convergence. Indeed, gain calibration reveals to be almost 19 times faster with StEFCal than with the WALS method described in [12]. Still, StEFCal was developed for estimation of direction independent gains, which is limited.

Considering all the specific introduced perturbations and their corresponding statistics enables to exploit all available information. In the following section, the Maximum A Posteriori (MAP) estimator is notably based on a priori knowledge.

2.4.2.4 Bayesian algorithm

Bayesian approaches are alternatives to ML-based techniques and one example is the MAP estimator [21,33], in which prior information is taken into account. As an example, the a priori model for the ionosphere should be accurate enough to reflect the fluctuations on a very small scale but not too complex so that the estimation algorithm can be deducted at low cost. In this sense, a statistical model can be introduced to describe the phase fluctuations within the ionosphere [21], notably high-order statistics in order to study local differences. Kolmogorov theory on turbulent environment is particularly well-adapted. Let us note that introducing many parameters to describe the model makes it more accurate and realistic, leading to small bias. On the other hand, with less parameters to estimate, we can hope to achieve a better variance, *i.e.*, a lower estimation error.

One of the most popular estimators in estimation theory is the Minimum Mean Squared Error (MMSE), given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \, \mathrm{E}\Big\{ ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2^2 \Big\}, \tag{2.40}$$

i.e., the minimization of the Mean Square Error (MSE). The corresponding solution is $\hat{\boldsymbol{\theta}} = \mathrm{E}\Big\{ \boldsymbol{\theta} | \mathbf{y} \Big\}$, so posterior mean of the parameters needs to be computed. According to Bayes's law, this distribution can be written as

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}}. \tag{2.41}$$

The MMSE is the optimal estimator in the LS sense but remains difficult to implement due to multidimensional integral and expected value from (2.41) and (2.40).

A good alternative is the MAP estimator which consists in maximizing the a posteriori distribution, *i.e.*,

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, p(\boldsymbol{\theta}|\mathbf{y}) = \arg\max_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}). \tag{2.42}$$

When the noise follows a Gaussian distribution, this amounts to solve a non-linear LS problem. In our interferometric calibration problem, the MAP estimator leads to [4,21]

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmax}} \, ||\boldsymbol{\Psi}^{-\frac{1}{2}} \left(\mathbf{y} - \mathbf{s}(\boldsymbol{\theta})\right)||_2^2 + ||\mathbf{C}_{\boldsymbol{\theta}}^{-\frac{1}{2}} \boldsymbol{\theta}||_2^2 \tag{2.43}$$

where $\boldsymbol{\Psi}$ is the covariance of the noise vector $\mathbf{n}$ and we note $\mathbf{C}_{\boldsymbol{\theta}}$ the covariance for $\boldsymbol{\theta}$. In the case of a pure Kolmogorov turbulent process, $\mathbf{C}_{\boldsymbol{\theta}}$ can be modeled by a power law with unknown parameters, thus, a priori knowledge is exploited and the additional term in (2.43) acts as a penalty function. Let us note that from a Bayesian perspective, the introduced hyperparameters in the model for $\mathbf{C}_{\boldsymbol{\theta}}$ must be estimated as they describe its statistical structure. However, the main issue is to obtain accurate models for the unknown perturbations, which are not always available in practice.

2.4.3 Variants to LS calibration

LS calibration represents the traditional estimator in case of additive white Gaussian noise. Still, there are some well-known disadvantages: speed of convergence is slow, convergence to global optimum is not guaranteed and the computational cost is significant. To improve both complexity and accuracy, alternative methods have been introduced.

2.4.3.1 EM algorithm

The EM algorithm [28, 100] is an iterative method which provides ML estimates asymptotically when observations are viewed as incomplete data, *i.e.*, some data points are unknown because they are missing or hidden. Compared to classical optimization algorithms used to solve non-linear LS problems, as the LM, the EM enables to reduce the computational cost and speed up the convergence.

It consists of two consecutive steps: an expectation (E) and a maximization (M) steps, and requires to define the so-called complete data $\mathbf{w}$, extracted from the output data $\mathbf{y}$. In order to specify the complete data, in our context, we assume well separated sources with unique propagation paths, from one source to one antenna, so that we can assign one unknown parameter vector $\boldsymbol{\theta}_i$ to each source for $i \in \{1, \dots, D\}$. Thus, a partitioning per source can be conducted. The case of sources sharing some parameters, due for instance to small angular deviation, is addressed in [89].

Instead of considering the complete data model with contributions from all directions, we reduce the problem to multiple sub-problems, leading to the following complete data vector for the i-th source [14, 89]

$$\mathbf{w}_i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n}_i \tag{2.44}$$

where the unknown parameter vector is decomposed as $\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \cdots, \boldsymbol{\theta}_D^T]^T$. The noise is also assumed to be decomposable into source contributions, *i.e.*, $\mathbf{n} = \sum_{i=1}^{D} \mathbf{n}_i$ and $\mathbf{y} = \sum_{i=1}^{D} \mathbf{w}_i$. To refine the description of the model, it is possible to associate more or less noise to each source, depending on its brightness, thanks to a weighting term.

The two steps of the EM algorithm are summarized as follows

• E step: in our case, it amounts to compute the expectation value of complete data conditionally to the observed data and the unknown parameter vector, *i.e.*,

$$\hat{\mathbf{w}} = \mathrm{E}\Big\{\mathbf{w}|\mathbf{y};\boldsymbol{\theta}\Big\}. \tag{2.45}$$

In the Gaussian case, a closed-form expression can be attained [101].

• M step: it consists in minimizing a cost function, for one given source i and repeat the estimation for $i \in \{1, \dots, D\}$ so the procedure is parallelisable. The cost function under study is thus given by $\zeta_i(\boldsymbol{\theta}_i) = ||\mathbf{w}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)||_F^2$, instead of $\zeta(\boldsymbol{\theta}) = ||\mathbf{y} - \sum_{i=1}^{D} \mathbf{s}_i(\boldsymbol{\theta})||_F^2$, which enables to reduce the complexity of the problem. To perform numerical minimization, the LM algorithm can still be used and computation is performed with respect to $\boldsymbol{\theta}_i$ instead of $\boldsymbol{\theta}$.

2.4.3.2 SAGE algorithm

The SAGE algorithm [102] is an alternative method to the EM. It can be applied for direction-dependent calibration of radio interferometers, leading to the SAGECal algorithm, whose implementation time and convergence speed are better than the EM. Compared to the latter, a different definition of the complete data is considered and the noise is assigned differently. Specifically, the hidden data are selected as [14]

$$\mathbf{w}_i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n} \tag{2.46}$$

s.t. $\mathbf{y} = \mathbf{w}_i + \sum\limits_{\substack{k=1 \\ k \neq i}}^{D} \mathbf{s}_k$. The two steps of the SAGE are quite similar to those of the EM: they consist in computing a conditional mean followed by a minimization process.

The SAGE algorithm is faster and leads to better calibration accuracy than the EM, but they are both superior in numerical complexity compared to direct application of the LM on the LS problem. To resolve issues on a more timely basis, it is therefore relevant to reduce the dimension of the unknown vector so as to invert matrices with lower dimensions. To ensure convergence towards a global optimum, probabilistic techniques as the Simulated Annealing (SA) [103] are possible but on the other hand, they unfortunately reduce the convergence speed.

The SAGE algorithm is particularly appropriate in some pathological cases when sources are not easily differentiable from the background noise or when they are close to each other [104] or even when positions of sensors are misspecified. Generally, classical calibration sources are viewed as bright and compact point sources with no broad or diffuse structure. But they may not be bright enough. To remedy this, we can introduce clustered sources [68] where each clustering acts as a single bright source. More information is available as more sources are considered all together and not individually. To this end, sources of the same cluster must be physically close to one another with a weighting to take into account their respective intensity and the perturbations must be similar due to small angular deviation. Such technique simplifies the problem since we reduce the number of directions to calibrate and we consider the information associated to several sources all together.

All the previously mentioned methods are applied assuming a Gaussian noise model for $\mathbf{n}_{pq}$ in (2.20) in the context of radio interferometry. Still, this is not always realistic and robustness needs to be addressed.

2.4.3.3 ECME algorithm

Due to the presence of outliers which corrupt the observations, the noise cannot simply be considered Gaussian [41]. These are due to radio frequency interferences which are not always perfectly filtered [15, 105, 106], to errors in the sky model due to weak unknown sources in the background [2], to the temporary appearance of particularly bright sources in the sidelobes generating punctual interferences [107].

To investigate robustness [108], the noise can be modeled as a Student's t distribution [109]. Such distribution is a suitable heavy-tailed statistical model for data sets affected by the presence of outliers. To provide relevant ML estimates with a Student's t noise model, the Expectation-Conditional Maximization Either (ECME) algorithm [110], an extension of the EM, is considered in [2, 41] due to its adequate rate of convergence.

Using (2.21), we distinguish the contribution of D known bright sources from the contribution of noise and D' unknown weak sources. The latter act as outliers and can be incorporated in a total non-Gaussian noise $\mathbf{n}$, *s.t.*,

$$\mathbf{y} = \sum_{s=1}^{D} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n} \tag{2.47}$$

in which $\mathbf{n} = \sum_{i'=1}^{D'} \mathbf{s}_{i'}(\boldsymbol{\theta}') + \mathbf{n}'$. Generally, $D' \gg D$ and only Jones matrices related to calibration sources, *i.e.*, $\boldsymbol{\theta}$, are estimated. Parameters $\boldsymbol{\theta}'$ can be deduced through interpolation for instance. By removing the contribution of D strong sources, the D' weakest contributions are then revealed if calibration has been performed accurately. One of the goals in calibration is to preserve the flux of the weakest sources in the resulting residual image [111].

Let us consider a Gaussian distribution $\mathcal{N}(\mu, \sigma_n^2/\gamma)$ where μ is the mean and the variance is given by a ratio. The denominator is a random variable following a Gamma distribution,

s.t., $\gamma \sim \text{Gamma}(\nu/2, \nu/2)$ with ν the number of degrees of freedom. The resulting marginal distribution is a Student's t defined by the following probability density function

$$\mathcal{ST}(t; \mu, \sigma_n^2, \gamma) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\pi\nu}\Gamma(\frac{\nu}{2})\sigma_n}\Big(1 + \frac{1}{\nu}\left(\frac{t-\mu}{\sigma_n}\right)^2\Big)^{-\frac{\nu+1}{2}}. \tag{2.48}$$

As the degree of freedom approaches infinity, the Student's t becomes a Gaussian distribution while it is heavy-tailed for low values of ν [112]. If $\mathbf{n}$ is assumed to follow the fixed distribution in (2.48) with $\sigma_n = 1$, $\mu = \mathbf{s}(\boldsymbol{\theta})$ and independent and identically distributed (i.i.d.) entries, the corresponding likelihood expression is a product between N components where N is the number of real data points in $\mathbf{y}$. In this specific case, maximizing the resulting likelihood does not provide the ML estimate as the noise is not Gaussian and ν needs to be estimated as well.

In the ECME, auxiliary variables are introduced, such as weights $\mathbf{w}$ and a scalar λ. Two steps are then considered:

- The E step consists in computing the following conditional expression

$$\hat{\mathbf{w}} = \text{E}\Big\{\mathbf{w}|\mathbf{y}; \boldsymbol{\theta}, \nu\Big\} \tag{2.49}$$

for each entry of the hidden vector $\mathbf{w}$. The scalar λ is also updated as function of the weights $\mathbf{w}$.

- The goal of the maximization step is to provide an estimate of ν by maximizing the log-likelihood expression *w.r.t.* ν and equate the result to zero. But the most important part is to obtain an estimate for $\boldsymbol{\theta}$. To this end, the following minimization process is considered

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\text{argmin}} \sum_{k=1}^{N} [\mathbf{w}]_k \Big([\mathbf{y}]_k - [\mathbf{s}(\boldsymbol{\theta})]_k\Big)^2 \tag{2.50}$$

with a weighted LS cost function. To solve (2.50), a robust version of the LM can be considered [41].

The robustness issue has been tackled but all the previously introduced algorithms consider one frequency bin at a time during calibration, which is not an optimal way to process the data. Therefore, in the following, we focus on a multi-frequency scenario considering the full observation bandwidth as a whole.

2.4.3.4 ADMM-based algorithm

To exploit frequency dependence of Jones matrices, it is possible to introduce polynomials, with a given order, assuming smooth variation across frequency. In a multi-frequency scenario, the computing load is particularly heavy due to large volume of data. To relieve this burden, data parallelism can be brought to bear by using a network of multiple agents among which data is distributed across frequency [35]. Therefore, data are not centralized but relocated in space and each agent has access to some specific data, at a given frequency. Calibration can then be reformulated as a consensus distributed optimization problem [36]. As a first step, we only consider frequency dependence here, in a given time interval.

For each frequency f, Jones matrices are modeled as follows to enforce smoothness

$$\mathbf{J}^{[f]} = \mathbf{P}^{[f]}\mathbf{Z} \tag{2.51}$$

where $\mathbf{P}^{[f]}$ collects all the frequency-dependent polynomial terms and $\mathbf{Z}$ refers to the hidden variables which need to be estimated.

Instead of estimating directly $\mathbf{Z}$ from all data available, distributed calibration consists in performing calibration locally by each agent, for a given subband. Thus, calibration amounts to solve the following constrained problem

$$\{\hat{\mathbf{J}}^{[f_1]}, \hat{\mathbf{J}}^{[f_2]}, \cdots, \hat{\mathbf{Z}}\} = \underset{\mathbf{J}^{[f_1]}, \mathbf{J}^{[f_2]}, \cdots, \mathbf{Z}}{\operatorname{argmin}} \sum_f \psi^{[f]}(\mathbf{J}^{[f]}) \quad \text{s.t.} \quad \mathbf{J}^{[f]} = \mathbf{P}^{[f]}\mathbf{Z} \tag{2.52}$$

in which $\psi^{[f]}(\cdot)$ is the following LS cost function

$$\psi^{[f]}(\boldsymbol{\theta}^{[f]}) = \sum_{pq} ||\mathbf{V}_{pq}^{[f]} - \sum_{i=1}^{D} \mathbf{J}_{i,p}^{[f]}(\boldsymbol{\theta}^{[f]}) \mathbf{C}_i^{[f]} \mathbf{J}_{i,q}^{[f]^H}(\boldsymbol{\theta}^{[f]})||_F^2. \tag{2.53}$$

To solve (2.52), we introduce unknown Lagrange multipliers $\mathbf{Y}^{[f]}$ and a fixed regularization term ρ leading to the following Lagrangian

$$\mathcal{L}(\mathbf{J}^{[f_1]}, \mathbf{J}^{[f_2]}, \cdots, \mathbf{Z}, \mathbf{Y}^{[f_1]}, \mathbf{Y}^{[f_2]}, \cdots) = \sum_f \mathcal{L}^{[f]}(\mathbf{J}^{[f]}, \mathbf{Z}, \mathbf{Y}^{[f]}) \tag{2.54}$$

in which

$$\mathcal{L}^{[f]}(\mathbf{J}^{[f]}, \mathbf{Z}, \mathbf{Y}^{[f]}) = \psi^{[f]}(\mathbf{J}^{[f]}) + ||\mathbf{Y}^{[f]^H} \left(\mathbf{J}^{[f]} - \mathbf{P}^{[f]}\mathbf{Z}\right)||_F + \frac{\rho}{2}||\mathbf{J}^{[f]} - \mathbf{P}^{[f]}\mathbf{Z}||_F^2. \tag{2.55}$$

A possible method to solve this problem is the Alternating Direction Method of Multipliers (ADMM) [36, 113, 114]. Instead of solving a global problem as $\operatorname{argmin} \sum_f \psi^{[f]}(\cdot)$ at the fusion center which collects data from all agents, calibration is performed at each agent locally, for one given frequency. Therefore, frequencies are processed independently but simultaneously and communication between computational agents is ensured thanks to the global fusion center. The study was originally introduced for one single source but it can easily be extended to the case of D calibration sources by using parameter partitioning with SAGE algorithm [89].

The three iterative steps of the ADMM are the following

$$\left(\hat{\mathbf{J}}^{[f]}\right)^{t+1} = \underset{\mathbf{J}^{[f]}}{\operatorname{argmin}}\, L^{[f]}\left(\boldsymbol{\theta}^{[f]}, \left(\hat{\mathbf{Z}}\right)^t, \left(\hat{\mathbf{Y}}^{[f]}\right)^t\right) \tag{2.56}$$

performed locally by each agent

$$\left(\hat{\mathbf{Z}}\right)^{t+1} = \underset{\mathbf{Z}}{\operatorname{argmin}} \sum_f L^{[f]}\left(\left(\hat{\mathbf{J}}^{[f]}\right)^{t+1}, \mathbf{Z}, \left(\hat{\mathbf{Y}}^{[f]}\right)^t\right) \tag{2.57}$$

performed at the fusion center

$$\left(\hat{\mathbf{Y}}^{[f]}\right)^{t+1} = \left(\hat{\mathbf{Y}}^{[f]}\right)^t + \rho\left(\left(\hat{\mathbf{J}}^{[f]}\right)^{t+1} - \mathbf{P}^{[f]}\left(\hat{\mathbf{Z}}\right)^{t+1}\right) \tag{2.58}$$

performed locally by each agent

where t is the iteration counter. Estimation in (2.56) can be done iteratively with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) [29] or the Riemannian Trust-Region (RTR) [115] algorithms.

Each Jones estimate $\hat{\mathbf{J}}^{[f]}$ obtained by each agent is transfered to the processing center which then estimates $\mathbf{Z}$ globally. Indeed, estimation of $\mathbf{Z}$ requires information across all frequencies and enforces consensus among all agents. Solving (2.57) leads to a closed-form expression. Finally, update of $\mathbf{Y}^{[f]}$ is performed in the last step by each corresponding agent. Let us note that similar techniques have been studied for image synthesis [116, 117].

The advantages of distributed optimization in the network include the reduced computational cost as calculus are distributed among several agents. To go even further, it would be meaningful to take into account temporal variation of Jones matrices or treat extreme cases, such as when the number of frequencies is very high compared to the number of available agents. But all these studies are still under investigation.

In the last section of this chapter, we address some notions as regards the imaging process, without going into details as this is not the central core of the manuscript.

2.5 Imaging in radio astronomy

Measured correlations for baselines (u, v) and true brightness image I are related by the Van Cittert-Zernike theorem in (2.5). Producing an image, namely map making, consists in inverting such expression and leads to the dirty image as in (2.7) due to the finite number of observations. The dirty image can also be written as a convolution between the desired image I and the Point Spread Function (PSF) B, also called the dirty beam, which is the inverse Fourier transform of the sampling function S, *i.e.*,

$$I_D = I * B \tag{2.59}$$

and

$$B(l, m) = \int\int S(u, v) e^{2\pi j(ul+vm)} du dv. \tag{2.60}$$

In order to recover the final restored image I from I_D, with high-resolution, a priori knowledge about B and deconvolution techniques are required. Thus, the deconvolution process is essentially a Fourier inversion problem and consists in removing the instrumental response. Let us note that autocorrelations are usually not used in the image formation, in order to reduce the influence of noise. Besides, forming an image with high dynamic range becomes all the more difficult to perform with the new generation of radio interferometers, the new challenges at stake and the significant perturbation effects which need to be calibrated and removed.

Many deconvolution algorithms are based on sequential source removal and the most popular one is CLEAN [118]: from a mostly empty sky model, composed of discrete point sources, it sequentially estimates the brightest source and subtracts its partial contribution from the dirty image using a so-called loop gain, until reaching a noise-like image where only the weakest possible sources remain. Iteratively, each source position and intensity are estimated and added to a point source list which will finally result in the global reconstructed image, convolved with an ideal reconstruction beam. Other alternative methods are possible, such as the Maximum Entropy Method (MEM) [119], Clark CLEAN and Cotton-Schwab (CS) which accelerate and reduce the computational load of CLEAN [45], the parametric estimation based algorithm Least Squares Minimum Variance Imaging (LS-MVI) [120] or even sparse reconstruction techniques [116, 121–123]. Besides, since the spectral dimension can no longer be omitted, three-dimensional reconstruction approaches are increasingly becoming prevalent [124]. As mentioned before, the traditional CLEAN is based on a point source model but extension with shapelets or wavelets is a possibility.

Lastly, let us note that two softwares have been implemented for calibration of the LOFAR [25,74] with direction dependent effects [10]: the MeqTrees software [37] which is able to simulate and calibrate radio astronomical measurements, and the Blackboard system [125, 126], named Black-Board Self-cal (BBS). MeqTrees is particularly flexible to specify the different perturbation effects to introduce and will be presented in Chapter 5 for realistic data sets.

Furthermore, traditional calibration and imaging were originally implemented in the Astronomical Image Processing System (AIPS) software package, which became AIPS++ [66] and is now referred to as CASA [49,50]. *Awimager* is an example of imager, on CASA. Contrary to non directional effects which are usually corrected separately from the imaging process, estimation of errors due to direction dependency is usually combined with the deconvolution step [63,65,69]. Therefore, reconstructed images will also be provided in this work to emphasize the benefit of the designed direction dependent calibration techniques.

2.6 Conclusion

Most calibration algorithms are LS based, which operate iteratively for one given frequency channel and reveal to be asymptotically efficient under a Gaussian noise model. In section 2.4.3.3, the Student's t distribution was specifically considered to include the contribution of interferences and weak background sources in radio interferometric data for non-Gaussian environment. In addition, the multi-frequency case was discussed in section 2.4.3.4 where the calibration task was reformulated as a consensus optimization problem. In the following chapters, we propose robust calibration algorithms based on a broader class of distributions so as to avoid model misspecifications and achieve more robustness. An extension to the multi-frequency scenario is also presented and realistic simulations are performed with the flexible software MeqTrees in order to reconstruct residual images.

Chapter 3

Robust calibration method in non-Gaussian environment

3.1 Motivations for robust calibration

Most calibration approaches are LS based [12, 13, 61, 95] and as mentioned in section 2.4.3, alternative algorithms as the EM [27, 100, 127] or the SAGE [102] have been proposed to enhance the convergence rate. Still, the Gaussian noise assumption is usually considered in radio astronomy [14,89], which is not always realistic and can lead to poor estimation performance. As explained in section 2.4.3.3, the presence of disruptive outliers has multiple causes: man-made radio signals generate frequency interferences which can be removed by flagging [128] but are not always perfectly filtered in practice. The solar activity or the punctual appearance of strong sources in the sidelobes constitute other sources of interference which can randomly generate outliers. Furthermore, radio interferometric data include the contribution of bright calibration sources but also of multiple unknown weak sources in the background leading to incomplete sky models with flux loss and spurious sources [111, 129]. Thus, the data model for baseline (p, q) can be written as

$$\mathbf{V}_{pq} = \sum_{i=1}^{D} \mathbf{J}_{i,p}(\boldsymbol{\theta}) \mathbf{C}_i \mathbf{J}_{i,q}^H(\boldsymbol{\theta}) + \mathbf{N}_{pq} \tag{3.1}$$

where the noise component $\mathbf{N}_{pq}$ includes the ambient noise $\mathbf{N}'_{pq}$, which is assumed Gaussian, but also the presence of all previously mentioned sources of interferences leading to outliers. Considering all available baselines leads to (2.47) with $\mathbf{n}'$ and $\mathbf{n}$, respectively, the total Gaussian and non-Gaussian noise components. The aim of robust calibration is thus to estimate $\boldsymbol{\theta}$ under non-Gaussian noise environment $\mathbf{n}$ [130, 131]. Let us note that during the imaging process, the D' weak unknown sources, hidden in $\mathbf{n}$, become the key elements of interest we wish to recover.

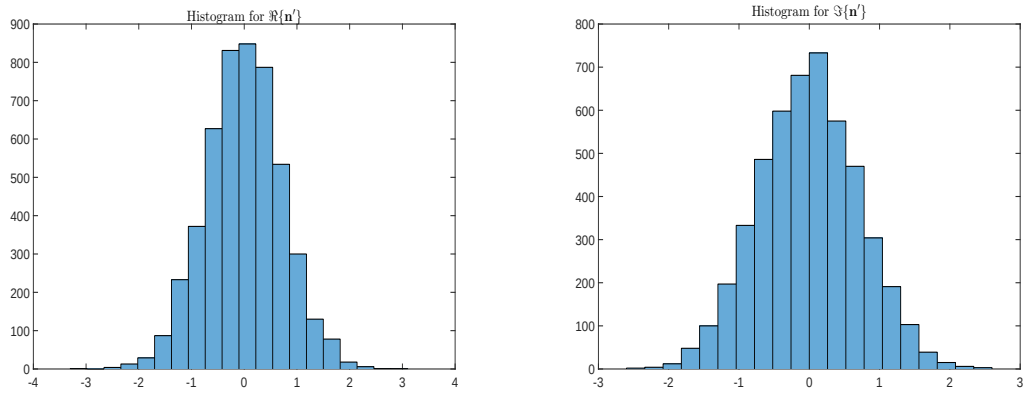
To robustify the calibration scheme, a distribution class, different from the Gaussian one, needs to be considered to model the noise contribution $\mathbf{n}$. To the best of our knowledge, the only robust calibration procedure was proposed in [2] and exposed in section 2.4.3.3. In this case, the noise part is specifically modeled as a Student's t distribution with i.i.d. entries, which is not optimal and leads to model misspecifications as the true distribution of $\mathbf{n}$ is not known in practice. In this chapter, we aim to show the limitations of the classical Gaussian noise assumption and propose an estimator based on a broader class of distributions.

3.1.1 Limits of the Gaussian noise modeling

The noise vector $\mathbf{n}_{pq}$ in (2.20) is traditionally assumed to follow a Gaussian distribution. However, in radio astronomy, this assumption may no longer be valid. In order to justify the use of non-Gaussian modeling in our application, we propose in the following some numerical studies about the statistical behavior of the noise in practical scenario.

To begin with, we generate radio interferometric data according to the model in (2.21). We consider $M = 8$ antennas, $D = 2$ slighly polarized calibration sources and the associated Jones matrices are randomly simulated as Gaussian complex numbers. We study the two following cases: i) the noise $\mathbf{n}'$ follows a typical Gaussian distribution and ii) D' unknown weak sources corrupt the observations as in (2.47). We assume that all sources, calibration and non-calibration ones, are affected by similar slowly varying directional errors and we choose $D' = 100$.

If we perform a test decision using the chi-square goodness-of-fit, numerical results indicate that the data in $\mathbf{n}'$ come indeed from a normal distribution while the realistic ones in $\mathbf{n}$ support the alternative hypothesis, *i.e.*, the data do not follow such distribution at the 1% significance level. This point is illustrated by Figures 3.1 and 3.2. We notice that the histograms for $\mathbf{n}$ are substantially different from those obtained with traditional Gaussian noise.



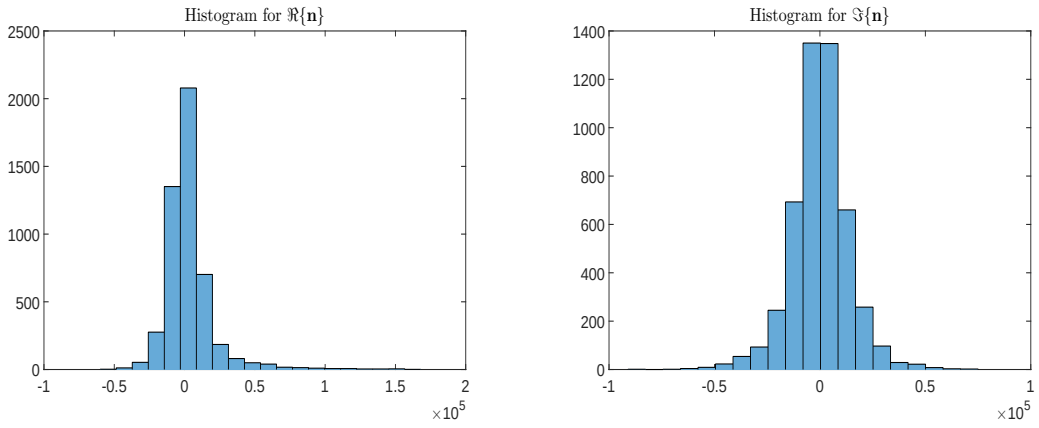
(a) (b)

Figure 3.1: Histograms for (a) $\Re\{\mathbf{n}'\}$ and (b) $\Im\{\mathbf{n}'\}$ (Gaussian case).

To confirm the non-Gaussianity of the noise, we plot in Figure 3.3 the histograms of the two different kinds of noisy data, along with a normal density function. We also display in Figure 3.4 the quantiles of sampled data vs. the theoretical quantiles from a standard normal distribution. We notice a linear straight line in the case of Gaussian noise $\mathbf{n}'$, which is consistent and deviations in the tails for $\mathbf{n}$ which illustrates its non-Gaussian nature. Therefore, the normal density is not so well fitted when contribution of D' weak unknown sources is taken into account. From the previous statistical studies, it seems clear that considering a Gaussian model for the noise in radio astronomy is not adapted.

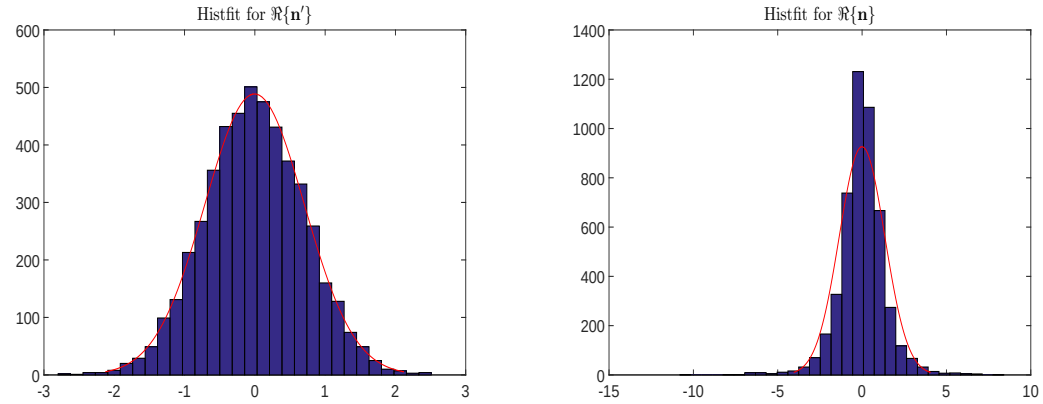
3.1.2 Compound-Gaussian noise modeling

To propose a suitable alternative to the typical Gaussian noise assumption in a robust scenario, we study a wide class of distributions, gathered under the so-called CG modeling [17,132]. Such distribution reveals to be suitable to achieve robustness *w.r.t.* outliers. Indeed, it encompasses a wide range of different noise distributions, among which the Student's t modeling



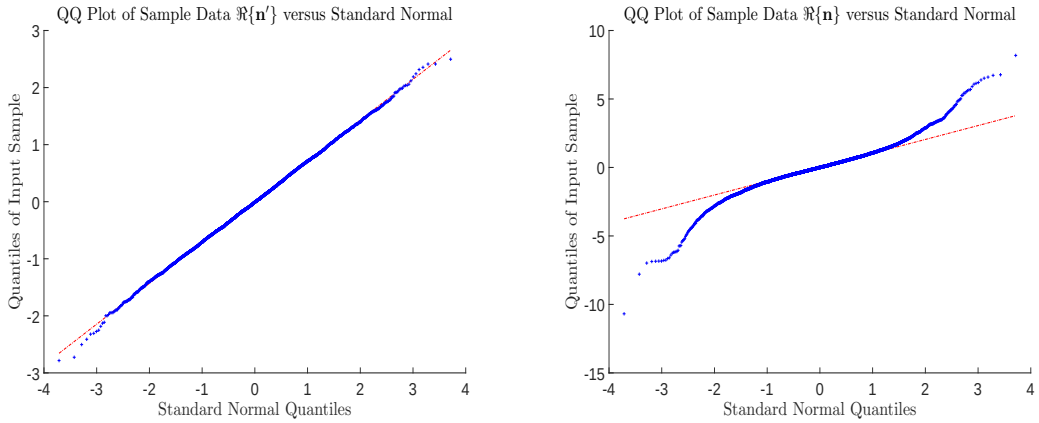
(a) (b)

Figure 3.2: Histograms for (a) $\Re\{\mathbf{n}\}$ and (b) $\Im\{\mathbf{n}\}$ (realistic case).



(a) (b)

Figure 3.3: Histograms with fitted normal distribution for (a) $\Re\{\mathbf{n}'\}$ and (b) $\Re\{\mathbf{n}\}$.



(a) (b)

Figure 3.4: Quantile-quantile plots for (a) $\Re\{\mathbf{n}'\}$ and (b) $\Re\{\mathbf{n}\}$.

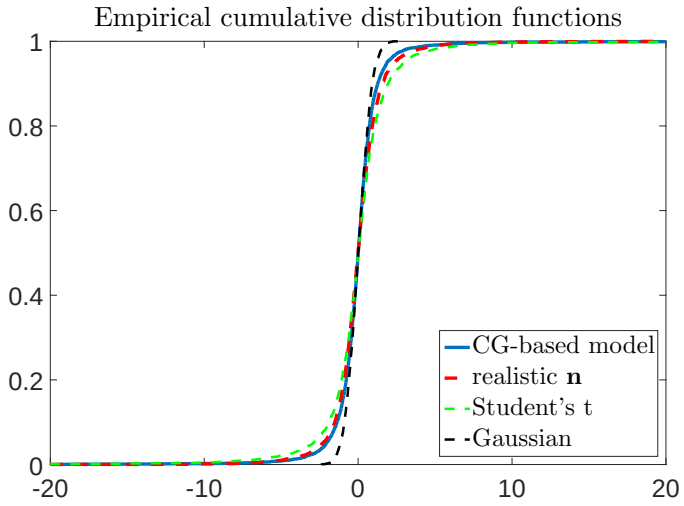


Figure 3.5: Empirical cumulative distribution plots for different distributions.

and the K-distribution. Its flexibility enables to adaptively consider non-Gaussian heavy-tailed density functions or the traditional Gaussian case when there are no outliers.

In (2.21), the noise includes contribution from all baselines, *i.e.*, $\mathbf{n} = [\mathbf{n}_{12}, \mathbf{n}_{13}, \ldots, \mathbf{n}_{(M-1)M}]^T$. For a particular antenna pair $(p, q) \in \{1, \ldots, M\}^2$, the two-scale CG distribution is generated as [133]

$$\mathbf{n}_{pq} = \sqrt{\tau_{pq}}\, \mathbf{g}_{pq}, \tag{3.2}$$

where τ_{pq} is a positive random variable, referred to as the texture part in the radar context, and $\mathbf{g}_{pq}$ is the speckle component, given by

$$\mathbf{g}_{pq} \sim \mathcal{CN}(\mathbf{0}, \mathbf{\Omega}) \tag{3.3}$$

which is a zero-mean Gaussian distribution with unknown 4×4 speckle covariance matrix $\mathbf{\Omega}$.

To consolidate our choice to consider such distribution to model the noise in visibility measurements, we compare in Figure 3.5 the empirical cumulative distribution functions of the generated data $\mathbf{n}$ with the typical Gaussian case, the Student's t distribution as chosen in [41] with fixed $\nu = 2$ here and the proposed two-scale CG modeling as introduced in (3.2). We notice that the nearest curve to the realistic noise part $\mathbf{n}$ is obtained with the CG-based model.

3.2 Model setup

We recall that the array is made up of M antennas with known locations and the sky is composed of D calibration sources. As exposed in section 2.3, each antenna provides sensitivity to both polarization directions (x, y) and the relation between the i-th incident electromagnetic wave and the generated voltage at the p-th antenna is given in (2.9).

3.2.1 Non-structured case

In the so-called non-structured case, each 2×2 Jones matrix $\mathbf{J}_{i,p}(\boldsymbol{\theta})$ is parametrized by the unknown vector $\boldsymbol{\theta}$ which corresponds to the entries of all Jones matrices [14, 134]. As we can associate a different Jones matrix with each signal propagation path i-p, the total number of

Jones matrices is DM and thus, $\boldsymbol{\theta}$ collects $8DM$ unknown real elements. We remind that Jones matrices model the array response and all environmental and instrumental distorsions along each propagation path.

In the radio interferometric systems under study, the measured observations consist of cross-correlations computed for each baseline, or antenna pair $(p,q) \in \{1,\ldots,M\}^2$, with $p < q$, resulting in (2.12), if we assume uncorrelated emitted signals. The 2×2 source coherency matrix $\mathbf{C}_i$ is assumed known from prior knowledge. Let us note that with the condition $p < q$, autocorrelations $\mathbf{V}_{pp}$ are ignored which is typical in radio astronomy as they are automically flagged [7].

Using the vectorization operator, in the noise-free case, leads to the following 4×1 vector

$$\tilde{\mathbf{v}}_{pq}(\boldsymbol{\theta}) = \mathrm{vec}\Big(\mathbf{J}_{i,p}(\boldsymbol{\theta})\mathbf{C}_i\mathbf{J}_{i,q}^H(\boldsymbol{\theta})\Big) = \sum_{i=1}^{D} \mathbf{s}_{i,pq}(\boldsymbol{\theta}) \tag{3.4}$$

where $\mathbf{s}_{i,pq}(\boldsymbol{\theta}) = \Big(\mathbf{J}_{i,q}^*(\boldsymbol{\theta}) \otimes \mathbf{J}_{i,p}(\boldsymbol{\theta})\Big)\,\mathbf{c}_i$ and $\mathbf{c}_i = \mathrm{vec}(\mathbf{C}_i)$. For noisy measurements, we introduce a noise vector $\mathbf{n}_{pq}$ for each antenna pair, leading to

$$\mathbf{v}_{pq} = \tilde{\mathbf{v}}_{pq}(\boldsymbol{\theta}) + \mathbf{n}_{pq} \tag{3.5}$$

which amounts to (2.20). By stacking all possible cross-correlations in a global $4B \times 1$ complex vector, we obtain the full data vector $\mathbf{y}$ in (2.21) where $\mathbf{s}_i(\boldsymbol{\theta}) = \Big[\mathbf{s}_{i,12}^T(\boldsymbol{\theta}), \mathbf{s}_{i,13}^T(\boldsymbol{\theta}), \ldots, \mathbf{s}_{i,(M-1)M}^T(\boldsymbol{\theta})\Big]^T$ and $\mathbf{n} = \Big[\mathbf{n}_{12}^T, \mathbf{n}_{13}^T, \ldots, \mathbf{n}_{(M-1)M}^T\Big]^T$ includes Gaussian background noise but also the eventual presence of outliers.

The non-structured case is particularly flexible as there is no need the specify the full propagation path, thus avoiding any model misspecifications, but the number of unknowns can be relatively large.

3.2.2 Structured case

In this section, we introduce specific models to describe the physical mechanism behind each perturbation effect: only physically meaningful parameters are estimated instead of all entries of all Jones matrices. To this end, we consider a particular context which is regime 3 of section 2.2.3 [3]: sensors are clustered into a compact array and their field-of-view are wide, resulting in direction dependent perturbations as propagation conditions differ depending on which source is targeted. But due to the small spatial extend, the same part of the ionosphere is probed and ionospheric distorsions are assumed similar for all receiving elements. Such calibration scenario is well-adapted for LOFAR or SKA stations [4,95].

In regime 3, a particular decomposition of each Jones matrix is given by [24–26,31]

$$\mathbf{J}_{i,p}(\boldsymbol{\theta}_{i,p}) = \mathbf{G}_p(\mathbf{g}_p)\mathbf{H}_{i,p}\mathbf{Z}_i(\varphi_i)\mathbf{F}_i(\vartheta_i) \tag{3.6}$$

for $i \in \{1,\ldots,D\}$, $p \in \{1,\ldots,M\}$ and $\boldsymbol{\theta}_{i,p} = [\vartheta_i, \varphi_i, \mathbf{g}_p^T]^T$.

• Ionospheric effects: As described in section 2.2.1, the ionosphere is a disturbed environment with spatial variations. One of the prevailing effects is a propagation delay which results in apparent shift of the source position [4,73]. We adapt (2.16) to our specific scenario, leading to

$$\mathbf{Z}_i(\varphi_i) = \exp\Big(j\varphi_i\Big)\mathbf{I}_2. \tag{3.7}$$

Propagation through the ionosphere also results in rotation effects whose form is given in (2.18). For the ionospheric Faraday rotation, we write it as

$$\mathbf{F}_i(\vartheta_i) = \begin{bmatrix} \cos(\vartheta_i) & -\sin(\vartheta_i) \\ \sin(\vartheta_i) & \cos(\vartheta_i) \end{bmatrix} \tag{3.8}$$

where ϑ_i is the unknown Faraday rotation angle.

As the array is compact with similar elements in regime 3, propagation delay (3.7) and rotation (3.8) are assumed identical for all antennas here [24].

• Instrumental effects: Each antenna is described by two complex electronic gains, *s.t.*,

$$\mathbf{G}_p(\mathbf{g}_p) = \text{diag}\{\mathbf{g}_p\} \tag{3.9}$$

where $\mathbf{g}_p$ is the 2×1 unknown complex gain vector.

Finally, $\mathbf{H}_{i,p}$ is an assumed known matrix for which prior information is provided by the available calibration source locations and antenna positions in (2.15) but also thanks to electromagnetic simulations as regards the antenna response and the beam pattern [18,24,31,32].

In this structured case, unknown parameters of interest are given by $\boldsymbol{\theta}_{i,p}$ for $i \in \{1, \ldots, D\}$ and $p \in \{1, \ldots, M\}$. We collect them all in $\boldsymbol{\varepsilon} = \mathbf{P}[\boldsymbol{\theta}_{1,1}^T, \boldsymbol{\theta}_{1,2}^T, \ldots, \boldsymbol{\theta}_{D,M}^T]^T$ where $\mathbf{P}$ is an appropriate rearrangement matrix, avoiding redundancy and leading to

$$\boldsymbol{\varepsilon} = [\vartheta_1, \ldots, \vartheta_D, \varphi_1, \ldots, \varphi_D, \mathbf{g}_1^T, \ldots, \mathbf{g}_M^T]^T. \tag{3.10}$$

Therefore, the global number of real unknowns is $2(D + M)$ and we assume $M \gg D$ so that observations outnumber unknown parameters.

3.3 Estimation of the texture and speckle parameters

Estimation of the noise parameters in (3.2) can be performed in two ways: either a priori knowledge is available about the distribution of τ_{pq} and a Bayesian approach is conducted or the distribution is unknown and it can be assumed deterministic during the estimation procedure. All the algorithms we are proposing in this work are iterative and based on the ML estimator [20,135]

3.3.1 Bayesian approach

Iterative MAP estimator (IMAPE): From a Bayesian perspective, we investigate robust calibration considering assumed specific distributions for the texture realizations τ_{pq}, resulting in MAP-based estimators [136]. Using Bayes' theorem, the expression of the joint log-likelihood function, denoted as $\mathcal{L}_J$, is given by

$$\mathcal{L}_J = \mathcal{L}_C + \sum_{pq} \ln\Big(p(\tau_{pq}; \boldsymbol{\varphi})\Big) \tag{3.11}$$

where $\mathcal{L}_C$ is the conditional likelihood function and $\boldsymbol{\varphi}$ refers to the unknown hyperparameters describing the corresponding prior distribution. Thus, such parameters need to be estimated as well. The principle of the MAP estimator is to maximize (3.11) *w.r.t.* each unknown individual parameter through a step-wise approach, leading to alternative update of unknowns $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\Omega}$ and $\boldsymbol{\tau}$. Assuming independence between $\mathbf{n}_{pq}$, the expression of $\mathcal{L}_C$ is given by

$$\mathcal{L}_C = \ln p\left(\mathbf{y}|\boldsymbol{\tau}; \boldsymbol{\theta}, \boldsymbol{\Omega}\right) = -\sum_{pq} \left(\frac{1}{\tau_{pq}} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}) + \ln|\pi \tau_{pq} \boldsymbol{\Omega}| \right) \tag{3.12}$$

in which $\boldsymbol{\tau} = [\tau_{12}, \tau_{13}, \ldots, \tau_{(M-1)M}]^T$ and $\mathbf{u}_{pq}(\boldsymbol{\theta}) = \mathbf{v}_{pq} - \sum_{i=1}^{D} \mathbf{s}_{i,pq}(\boldsymbol{\theta})$.

As it can been seen from (3.2), the probability density function (pdf) of each texture parameter τ_{pq} needs to be specified. Depending on the nature of τ_{pq}, different algorithms can be developed. We exploit prior distributions for the texture, available in closed-form, which generate various heavy-tailed noise models for $\mathbf{n}_{pq}$. In the following, we define specific texture priors and derive the corresponding IMAPE for each prior distribution $p(\tau_{pq}; \boldsymbol{\varphi})$.

• Gamma distribution: The corresponding pdf is given by

$$p(\tau_{pq}; a, b) = \frac{1}{\Gamma(a) b^a} \tau_{pq}^{a-1} \exp\left(-\frac{\tau_{pq}}{b}\right) \tag{3.13}$$

where a and b are the shape and scale parameters, respectively. Such prior leads to the K-distribution for the noise vector $\mathbf{n}_{pq}$ [16].

If we insert (3.13) into (3.11), the resulting joint log-likelihood function reads

$$\mathcal{L}_J = \mathcal{L}_C + (a-1) \sum_{pq} \ln(\tau_{pq}) - \frac{\sum_{pq} \tau_{pq}}{b} - B \ln\Big(\Gamma(a)\Big) - Ba \ln(b).$$

Considering $\partial \mathcal{L}_J / \tau_{pq} = 0$ leads to the following estimate for the texture parameter

$$\hat{\tau}_{pq} = \frac{(a-5)b + \Big((a-5)^2 b^2 + 4b \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta})\Big)^{1/2}}{2} \tag{3.14}$$

while $\partial \mathcal{L}_J / \boldsymbol{\varphi} = 0$ results in

$$\hat{b} = \frac{\sum_{pq} \tau_{pq}}{Ba} \tag{3.15}$$

for the scale, and the following equation to solve for the shape

$$-B\Psi(\hat{a}) + \sum_{pq} \ln(\tau_{pq}) - B \ln(b) = 0 \tag{3.16}$$

which can be computed numerically.

In this specific case, steps of the corresponding IMAPE are described in Algorithm 1. Due to the iterative approach, we notice that (3.15) was plugged into (3.16) as well as (3.14) into (3.28), see section 3.3.2. Similar algorithms can be derived for each texture prior.

• Inverse Gamma distribution: When the texture parameter follows an inverse Gamma function, *s.t.*,

$$p(\tau_{pq}; a, b) = \frac{b^a}{\Gamma(a)} \tau_{pq}^{-a-1} \exp\left(-\frac{b}{\tau_{pq}}\right), \tag{3.17}$$

the generated noise model is a Student's t distribution [109, 137]. Let us note that when the shape parameter is fixed in (3.17), *s.t.*, $a = 1$, the CG-based model for $\mathbf{n}_{pq}$ becomes a Cauchy distribution [17].

Using (3.17), (3.11) and the derivative *w.r.t.* τ_{pq}, the corresponding estimate reads

$$\hat{\tau}_{pq} = \frac{b + \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta})}{a+5}. \tag{3.18}$$

Algorithm 1: IMAPE Iterative MAP estimator / K-distributed noise

input : D, M, B, $\{\mathbf{C}_i\}_{i\in\{1,\dots,D\}}$, $\mathbf{y}$
output : $\hat{\boldsymbol{\theta}}$
initialize: $\hat{\boldsymbol{\Omega}} \leftarrow \boldsymbol{\Omega}_{\text{init}}$, $\hat{\boldsymbol{\tau}} \leftarrow \boldsymbol{\tau}_{\text{init}}$, $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_{\text{init}}$
while *stop criterion unreached* **do**
1. Obtain $\hat{\boldsymbol{\theta}}$, see section 3.4
2. Obtain $\hat{a}$ from solving $-B\Psi(a) + \sum_{pq} \ln(\hat{\tau}_{pq}) - B\ln\left(\frac{\sum_{pq}\hat{\tau}_{pq}}{Ba}\right) = 0$
3. Obtain $\hat{b}$ from (3.15)
4. Obtain $\hat{\boldsymbol{\Omega}}$ from $\hat{\boldsymbol{\Omega}} = \frac{2}{B}\sum_{pq} \frac{\mathbf{u}_{pq}(\hat{\boldsymbol{\theta}})\mathbf{u}_{pq}^H(\hat{\boldsymbol{\theta}})}{(\hat{a}-5)\hat{b}+\left((\hat{a}-5)^2\hat{b}^2+4\hat{b}\mathbf{u}_{pq}^H(\hat{\boldsymbol{\theta}})\hat{\boldsymbol{\Omega}}^{-1}\mathbf{u}_{pq}(\hat{\boldsymbol{\theta}})\right)^{1/2}}$ and (3.30)
5. Obtain $\hat{\boldsymbol{\tau}}$ from (3.14)

The scale is given by the following analytical expression

$$\hat{b} = \frac{Ba}{\sum_{pq}\frac{1}{\tau_{pq}}} \tag{3.19}$$

and the shape requires to solve

$$-B\Psi(a) - \sum_{pq}\ln(\tau_{pq}) + B\ln(b) = 0. \tag{3.20}$$

We mentioned the particular case of Cauchy distribution for which (3.20) becomes $\hat{a} = 1$.

• Exponential distribution: Such prior for the texture reads

$$p(\tau_{pq};\lambda) = \lambda\exp\left(-\lambda\tau_{pq}\right), \tag{3.21}$$

where λ is the so-called rate parameter. The resulting model for the noise is a Laplace distribution [17].

Based on (3.21) and (3.11), the calculations result in

$$\hat{\tau}_{pq} = \frac{-4 + \left(16 + 4\lambda\mathbf{u}_{pq}^H(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\mathbf{u}_{pq}(\boldsymbol{\theta})\right)^{1/2}}{2\lambda} \tag{3.22}$$

and the single rate parameter is deducted from

$$\hat{\lambda} = \frac{B}{\sum_{pq}\tau_{pq}}. \tag{3.23}$$

• Inverse Gaussian distribution: The inverse Gaussian distribution, also called Wald distribution, with shape λ and assumed unit mean [16], is described by the following pdf:

$$p(\tau_{pq};\lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2}\tau_{pq}^{-3/2}\exp\left(\frac{-\lambda(\tau_{pq}-1)^2}{2\tau_{pq}}\right). \tag{3.24}$$

The resulting CG model is referred to as the IG-CG distribution.

Estimates of noise parameters are given by

$$\hat{\tau}_{pq} = \frac{-11 + \left(121 + 4\lambda\left(\lambda + 2\mathbf{u}_{pq}^H(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\mathbf{u}_{pq}(\boldsymbol{\theta})\right)\right)^{1/2}}{2\lambda} \tag{3.25}$$

and

$$\hat{\lambda} = \frac{B}{\sum\limits_{pq} \frac{(\tau_{pq}-1)^2}{\tau_{pq}}}. \tag{3.26}$$

To sum up, we expose in Table 3.1 the different kinds of CG-based noise models we are considering in this study and their corresponding texture priors with estimates.

Table 3.1: Overview of the different MAP-based estimators.

Texture distribution	Expression of texture pdf	Noise pdf	$\hat{\tau}_{pq}$	$\hat{\boldsymbol{\varphi}}$
Gamma	$p(\tau_{pq};a,b) = \frac{1}{\Gamma(a)b^a}\tau_{pq}^{a-1}\exp\left(-\frac{\tau_{pq}}{b}\right)$	K	(3.14)	(3.16) & (3.15)
Inverse Gamma	$p(\tau_{pq};a,b) = \frac{b^a}{\Gamma(a)}\tau_{pq}^{-a-1}\exp\left(-\frac{b}{\tau_{pq}}\right)$,	Student's t	(3.18)	(3.20) & (3.19)
Inverse Gamma	$p(\tau_{pq};b) = b\tau_{pq}^{-2}\exp\left(-\frac{b}{\tau_{pq}}\right)$,	Cauchy	(3.18)	$a=1$ & (3.19)
Exponential	$p(\tau_{pq};\lambda) = \lambda\exp\left(-\lambda\tau_{pq}\right)$	Laplace	(3.22)	(3.23)
Inverse Gaussian	$p(\tau_{pq};\lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2}\tau_{pq}^{-3/2}\exp\left(\frac{-\lambda(\tau_{pq}-1)^2}{2\tau_{pq}}\right)$	IG-CG	(3.25)	(3.26)

3.3.2 (Relaxed) deterministic approach

The noise model is not exactly known in practice, so prior knowledge about the pdf is not always available. Thus, the statistical distribution of the texture parameter being unknown, we can consider it as an unknown deterministic parameter in the estimation process [22,23]. This avoids to choose a specific distribution which may lead to model misspecifications and enables to propose a broad robust estimator *w.r.t.* the presence of outliers. We call it the relaxed deterministic approach as it is not strictly based on the exact model in (3.2) [138,139].

Relaxed concentrated ML estimator (RCMLE): Optimization of (3.12) is performed *w.r.t.* each unknown parameter sequentially, while the others are fixed. Closed-form expressions can be obtained for the texture component τ_{pq}, for which no distribution has to be specified, and the speckle covariance matrix $\boldsymbol{\Omega}$. Estimation of $\boldsymbol{\theta}$ is the step of interest and will be discussed in section 3.4. If we take the derivative of (3.12) *w.r.t.* τ_{pq} and equate it to 0, we obtain the following expression for the texture estimate

$$\hat{\tau}_{pq} = \frac{1}{4}\mathbf{u}_{pq}^H(\boldsymbol{\theta})\boldsymbol{\Omega}^{-1}\mathbf{u}_{pq}(\boldsymbol{\theta}). \tag{3.27}$$

Using classical differential properties [140, p. 2741] and the permutation property, estimation of the speckle covariance matrix reads

$$\hat{\boldsymbol{\Omega}} = \frac{1}{B}\sum_{pq}\frac{1}{\tau_{pq}}\mathbf{u}_{pq}(\boldsymbol{\theta})\mathbf{u}_{pq}^H(\boldsymbol{\theta}). \tag{3.28}$$

We adopt an iterative procedure here with a concentrated ML scheme. Thus, by plugging (3.27) into (3.28), we finally obtain

$$\hat{\boldsymbol{\Omega}}^{t+1} = \frac{4}{B}\sum_{pq}\frac{\mathbf{u}_{pq}(\boldsymbol{\theta})\mathbf{u}_{pq}^H(\boldsymbol{\theta})}{\mathbf{u}_{pq}^H(\boldsymbol{\theta})\left(\hat{\boldsymbol{\Omega}}^t\right)^{-1}\mathbf{u}_{pq}(\boldsymbol{\theta})} \tag{3.29}$$

with t the iteration counter. Let us note that a constraint is required to remove scaling ambiguities in model (3.2). Therefore, we impose $\mathrm{tr}\{\boldsymbol{\Omega}\} = 1$ but this choice is arbitrary and does not affect the estimates of interest [20]. The normalization requires the following step in the procedure

$$\hat{\boldsymbol{\Omega}}^{t+1} = \frac{\hat{\boldsymbol{\Omega}}^{t+1}}{\mathrm{tr}\left\{\hat{\boldsymbol{\Omega}}^{t+1}\right\}}. \tag{3.30}$$

The global scheme of the proposed RCMLE is described in Algorithm 2.

Algorithm 2: RCMLE Relaxed concentrated ML estimator

input : D, M, B, $\{\mathbf{C}_i\}_{i\in\{1,\ldots,D\}}$, $\mathbf{y}$
output : $\hat{\boldsymbol{\theta}}$
initialize: $\hat{\boldsymbol{\Omega}} \leftarrow \boldsymbol{\Omega}_{\text{init}}$, $\hat{\boldsymbol{\tau}} \leftarrow \boldsymbol{\tau}_{\text{init}}$, $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_{\text{init}}$
while *stop criterion unreached* **do**
1 Obtain $\hat{\boldsymbol{\theta}}$, see section 3.4
2 Obtain $\hat{\boldsymbol{\Omega}}$ from (3.29) and (3.30),
3 Obtain $\hat{\boldsymbol{\tau}}$ from (3.27)

Remark: Let us note that it is possible to consider a different covariance matrix $\boldsymbol{\Omega}_{pq}$ for each baseline (p,q) in (3.3). In this specific case, the proposed robust calibration algorithm requires a few modifications and corresponding expressions are presented in Appendix A.

3.4 Estimation of Jones matrices

The proposed robust calibration estimator is based on the CG model in (3.2). We assume that noise parameters τ_{pq} and $\boldsymbol{\Omega}$ are estimated with the Bayesian or deterministic approach. For sake of clarity, we consider the relaxed concentrated ML method for which texture realizations are assumed deterministic. But the corresponding expressions for the Bayesian case are straightforward to obtain, using section 3.3.1.

Specifically, each block of unknown parameters is updated sequentially in a global iterative loop as shown in Algorithm 2. Estimates of noise parameters, *i.e.*, the texture and the speckle parts, are derived from (3.27) and (3.29), respectively, followed by the normalization constraint in (3.30), for identifiability issues.

3.4.1 Non-structured case

The goal is to estimate $\boldsymbol{\theta}$, considering (3.12). For a given $\hat{\boldsymbol{\Omega}}$ and $\hat{\boldsymbol{\tau}}$, it consists in the following minimization problem

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\mathrm{argmin}} \left\{ \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \boldsymbol{\Omega}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}) \right\}. \tag{3.31}$$

Thus, a multi-dimensional optimization approach needs to be carried out to solve (3.31), resulting in significant computation time due to the large size of $\boldsymbol{\theta}$. To reduce the computational cost, we propose to use the EM algorithm [27, 100, 127] which requires a proper parameter partitioning and the definition of complete data. In the non-structured case, it is natural to decompose the parameter of interest $\boldsymbol{\theta}$ as follows

$$\boldsymbol{\theta} = [\boldsymbol{\theta}_1^T, \ldots, \boldsymbol{\theta}_D^T]^T = [\boldsymbol{\theta}_{1,1}^T, \ldots, \boldsymbol{\theta}_{1,M}^T, \ldots, \boldsymbol{\theta}_{D,1}^T, \ldots, \boldsymbol{\theta}_{D,M}^T]^T. \tag{3.32}$$

This way, each propagation path *i-p* is parametrized by the vector $\boldsymbol{\theta}_{i,p} \in \mathbb{R}^{8\times 1}$, *s.t.*, $\mathbf{J}_{i,p}(\boldsymbol{\theta}) = \mathbf{J}_{i,p}(\boldsymbol{\theta}_{i,p})$.

As explained in section 2.4.3.1, the EM algorithm iterates between two steps and provides the ML estimates with limited complexity. These two steps are explained in the following

• E-step: The so-called complete data vector is given by (2.44) with a decomposition into source contributions *s.t.* $\mathbf{y} = \sum_{i=1}^{D} \mathbf{w}_i$ and $\mathbf{n} = \sum_{i=1}^{D} \mathbf{n}_i$. Each noise part is generated as

$$\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \beta_i \boldsymbol{\Psi}) \tag{3.33}$$

with $\sum_{i=1}^{D} \beta_i = 1$ and we note $\boldsymbol{\Psi}$ the $4B \times 4B$ covariance matrix of $\mathbf{n}$. With (3.2) and (3.3), we can write

$$\mathbf{n}_{pq} \sim \mathcal{CN}(\mathbf{0}, \tau_{pq}\boldsymbol{\Omega}). \tag{3.34}$$

Since measurements are assumed independent between antenna pairs, as seen in (3.12), the covariance $\boldsymbol{\Psi}$ has the following block-diagonal expression

$$\boldsymbol{\Psi} = \text{bdiag}\{\tau_{12}\boldsymbol{\Omega}, \ldots, \tau_{(M-1)M}\boldsymbol{\Omega}\}. \tag{3.35}$$

The complete data vector $\mathbf{w} = [\mathbf{w}_1^T, \ldots, \mathbf{w}_D^T]^T$ is defined by the following $4DB \times 4DB$ covariance matrix

$$\boldsymbol{\Xi} = \text{bdiag}\{\beta_1\boldsymbol{\Psi}, \ldots, \beta_D\boldsymbol{\Psi}\}. \tag{3.36}$$

Considering (2.45) for each source $i \in \{1, \ldots, D\}$ and [101, p. 36], the conditional expectation reads

$$\begin{aligned} \hat{\mathbf{w}}_i &= \text{E}\Big\{\mathbf{w}_i|\mathbf{y}; \boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega}\Big\} \\ &= \mathbf{s}_i(\boldsymbol{\theta}_i) + \beta_i \left(\mathbf{y} - \sum_{l=1}^{D} \mathbf{s}_l(\boldsymbol{\theta}_l)\right). \end{aligned} \tag{3.37}$$

• M-step: Instead of directly estimating $\boldsymbol{\theta} \in \mathbb{C}^{4DM\times 1}$ from (3.31), which can be computationally prohibitive, the global multiple source estimation problem is reduced to multiple single source sub-problems. Indeed, once $\hat{\mathbf{w}}$ is evaluated by the previous E-step, we need to maximize the following likelihood function

$$\begin{aligned} p(\hat{\mathbf{w}}|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\Omega}) &= \frac{1}{|\pi\boldsymbol{\Xi}|} \exp\left\{-\Big(\hat{\mathbf{w}} - \mathbf{s}(\boldsymbol{\theta})\Big)^H \boldsymbol{\Xi}^{-1}\Big(\hat{\mathbf{w}} - \mathbf{s}(\boldsymbol{\theta})\Big)\right\} \\ &= \prod_{i=1}^{D} \frac{1}{|\pi\beta_i\boldsymbol{\Psi}|} \exp\left\{-\Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big)^H (\beta_i\boldsymbol{\Psi})^{-1}\Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big)\right\} \end{aligned} \tag{3.38}$$

with assumed independent $\mathbf{w}_i$. Therefore, estimation of $\boldsymbol{\theta}_i \in \mathbb{C}^{4M\times 1}$ is obtained through minimization of the following objective function

$$\phi_i(\boldsymbol{\theta}_i) = \Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big)^H (\beta_i\boldsymbol{\Psi})^{-1}\Big(\hat{\mathbf{w}}_i - \mathbf{s}_i(\boldsymbol{\theta}_i)\Big). \tag{3.39}$$

This can be performed numerically with the LM algorithm [29, 88, 90], as exposed in section 2.4.2.1. However, obtaining closed-form expressions enables to reduce even more the complexity of the problem. To this end, we apply the BCD algorithm [30, 141] for each single source sub-problems in (3.39).

As shown in (3.32), the parameter vector is partitioned according to each source but also to each antenna. Thus, in the following, instead of optimizing (3.39) *w.r.t.* $\boldsymbol{\theta}_i$, we propose to carry out the optimization *w.r.t.* $\boldsymbol{\theta}_{i,p}$, for fixed $\boldsymbol{\theta}_{i,q}$ with $q \neq p$. We will see that an analytical solution is obtained for $\hat{\boldsymbol{\theta}}_{i,p}$ as function of all other $\boldsymbol{\theta}_{i,q}$ with $q \neq p$. The procedure is repeated alternatively for each component vector $\boldsymbol{\theta}_{i,p}$ with $p \in \{1, \ldots, M\}$ until convergence.

In (3.39), only $\mathbf{s}_i$ is function of $\boldsymbol{\theta}_i$ but actually, only a subset depends on the block-coordinate vector $\boldsymbol{\theta}_{i,p}$, *i.e.*, $\{\mathbf{s}_{i,pq}\}$ for $q > p$, $q \in \{1, \ldots, M\}$ and $\{\mathbf{s}_{i,qp}\}$ for $q < p$, $q \in \{1, \ldots, M\}$. Thus, it is more convenient to rewrite (3.39) as follows

$$\begin{aligned}\phi_i(\boldsymbol{\theta}_{i,p}) = &\sum_{\substack{q=1\\q>p}}^{M} \Big(\mathbf{w}_{i,pq} - \mathbf{s}_{i,pq}(\boldsymbol{\theta}_{i,p})\Big)^H (\beta_i \tau_{pq} \boldsymbol{\Omega})^{-1} \Big(\mathbf{w}_{i,pq} - \mathbf{s}_{i,pq}(\boldsymbol{\theta}_{i,p})\Big) + \\ &\sum_{\substack{q=1\\q<p}}^{M} \Big(\mathbf{w}_{i,qp} - \mathbf{s}_{i,qp}(\boldsymbol{\theta}_{i,p})\Big)^H (\beta_i \tau_{qp} \boldsymbol{\Omega})^{-1} \Big(\mathbf{w}_{i,qp} - \mathbf{s}_{i,qp}(\boldsymbol{\theta}_{i,p})\Big) + \text{Constant}. \end{aligned} \tag{3.40}$$

In the Constant part, we incorporate all subsets of $\phi_i(\boldsymbol{\theta}_{i,p})$ which do not depend on $\boldsymbol{\theta}_{i,p}$. Notation, calculus and details are given in Appendix B in which we show that it is possible to write

$$\mathbf{s}_{i,pq}(\boldsymbol{\theta}_{i,p}) = \boldsymbol{\Sigma}_{i,q} \boldsymbol{\theta}_{i,p} \tag{3.41}$$

and, likewise

$$\mathbf{s}_{i,qp}(\boldsymbol{\theta}_{i,p}) = \boldsymbol{\Upsilon}_{i,q} \boldsymbol{\theta}^*_{i,p}. \tag{3.42}$$

In the end, the expression of $\boldsymbol{\theta}_{i,p}$ for $p \in \{1, \ldots, M\}$ and the given i-th source direction is written as

$$\hat{\boldsymbol{\theta}}_{i,p} = \begin{cases} (\boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \boldsymbol{\Sigma}_i + \boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \boldsymbol{\Upsilon}_i)^{-1} (\boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \mathbf{w}_{i,p} + \boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \tilde{\mathbf{w}}_{i,p}) & \text{for } 1 < p < M \\ (\boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \boldsymbol{\Sigma}_i)^{-1} \boldsymbol{\Sigma}_i^H \mathbf{A}_{i,p} \mathbf{w}_{i,p} & \text{for } p = 1 \\ (\boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \boldsymbol{\Upsilon}_i)^{-1} \boldsymbol{\Upsilon}_i^H \tilde{\mathbf{A}}_{i,p} \tilde{\mathbf{w}}_{i,p} & \text{for } p = M \end{cases} \tag{3.43}$$

Finally, from the above discussion, the detailed scheme of the proposed RCMLE is summarized in Algorithm 3.

Algorithm 3: RCMLE Relaxed concentrated ML estimator

input : D, M, B, $\{\mathbf{C}_i, \beta_i\}_{i \in \{1,\ldots,D\}}$, $\mathbf{y}$
output : $\hat{\boldsymbol{\theta}}$
initialize: $\hat{\boldsymbol{\Omega}} \leftarrow \boldsymbol{\Omega}_{\text{init}}$, $\hat{\boldsymbol{\tau}} \leftarrow \boldsymbol{\tau}_{\text{init}}$, $\hat{\boldsymbol{\theta}} \leftarrow \boldsymbol{\theta}_{\text{init}}$
while *stop criterion unreached (**referred to as the first loop**)* **do**
  while *stop criterion unreached (**referred to as the second loop**)* **do**
1 E-step: $\hat{\mathbf{w}}_i$ obtained from (3.37), $i \in \{1, \ldots, D\}$
2 M-step: $\hat{\boldsymbol{\theta}}_i$ obtained as follows, $i \in \{1, \ldots, D\}$
    while *stop criterion unreached (**referred to as the third loop**)* **do**
3 $\hat{\boldsymbol{\theta}}_{i,p}$ obtained from (3.43), $p \in \{1, \ldots, M\}$
4 Obtain $\hat{\boldsymbol{\Omega}}$ from (3.29) and (3.30) ,
5 Obtain $\hat{\boldsymbol{\tau}}$ from (3.27)

Regarding the convergence properties of the proposed algorithm, we notice that: the maximization step of the EM algorithm is computed thanks to the BCD algorithm by considering

the cost function in (3.39). Minimization leads to a unique solution for each individual block of variables, given in (3.43). Thus, convergence to a stationary point and local convergence are ensured [142]. When the M-step is solved exactly, *i.e.*, when the BCD algorithm provides the exact minimizer of (3.39), and for a theoretical infinite number of iterations, convergence to a stationary point by the EM algorithm is ensured. Still, let us note that depending on the initialization, the stationary point can be either a global, a local minimizer or, in some unusual cases, a saddle point [127]. Proper initialization is thus required based on possible prior information or one can use some metaheuristic approaches such as random-restart hill climbing or SA methods. If the EM provides exact solutions, monotone local convergence of the ML estimator is ensured since the value of the cost function at each step can either improve or maintain but cannot worsen [143]. Therefore, convergence to a local stationary point is ensured for the overall algorithm if convergence is attained in each loop (three loops as shown in Algorithm 3), which requires an adequate initialization and a theoretical infinite number of iterations. However, in practice, numerical simulations in section 3.5 show that a few iterations are enough to reach convergence to, at least, a local minimizer and achieve relatively good numerical stability.

3.4.2 Structured case

In the structured case, we consider the specific model in (3.6), adapted to regime 3. The RCMLE in Algorithm 3 provides an estimate for all $\hat{\mathbf{J}}_{i,p}$ with $i \in \{1,\dots,D\}$ and $p \in \{1,\dots,M\}$. The idea is to estimate the unknown parameter vector of interest $\boldsymbol{\varepsilon}$ in a sequential manner, once RCMLE has been performed and all Jones entries have been estimated. To this end, we adopt an iterative procedure and optimize a given cost function *w.r.t.* one of the physical parameters in $\boldsymbol{\varepsilon}$ while fixing the others, leading to a global alternative method.

- Estimation of $\mathbf{g}_p$: Electronic gains are obtained by solving

$$\hat{\mathbf{g}}_p = \underset{\mathbf{g}_p}{\operatorname{argmin}}\, \kappa(\mathbf{g}_p) \tag{3.44}$$

in which

$$\begin{aligned}\kappa(\mathbf{g}_p) &= \sum_{i=1}^{D} ||\hat{\mathbf{J}}_{i,p} - \mathbf{G}_p(\mathbf{g}_p)\mathbf{H}_{i,p}\mathbf{Z}_i\mathbf{F}_i||_F^2 \\ &= \sum_{i=1}^{D} \mathrm{Tr}\Big\{ \Big(\hat{\mathbf{J}}_{i,p} - \mathbf{G}_p(\mathbf{g}_p)\mathbf{R}_{i,p}\Big)\Big(\hat{\mathbf{J}}_{i,p} - \mathbf{G}_p(\mathbf{g}_p)\mathbf{R}_{i,p}\Big)^H \Big\}\end{aligned} \tag{3.45}$$

where $\mathbf{R}_{i,p} = \mathbf{H}_{i,p}\mathbf{Z}_i\mathbf{F}_i$. Taking the derivative *w.r.t.* $[\mathbf{g}_p]_k$ for $k \in \{1,2\}$ leads to

$$\frac{\partial \kappa(\mathbf{g}_p)}{\partial [\mathbf{g}_p]_k} = \sum_{i=1}^{D} \mathrm{Tr}\Big\{ -\mathbf{e}_k\mathbf{e}_k^T\mathbf{R}_{i,p}\hat{\mathbf{J}}_{i,p}^H + \mathbf{e}_k\mathbf{e}_k^T\mathbf{R}_{i,p}\mathbf{R}_{i,p}^H\mathbf{G}_p^H \Big\}. \tag{3.46}$$

Equating (3.46) to 0 results in

$$\sum_{i=1}^{D}[\mathbf{X}_{i,p}]_{k,k} = \sum_{i=1}^{D}[\mathbf{W}_{i,p}\hat{\mathbf{G}}_p^H]_{k,k} = \sum_{i=1}^{D}[\mathbf{W}_{i,p}]_{k,k}[\hat{\mathbf{G}}_p^*]_{k,k} \tag{3.47}$$

where $\mathbf{X}_{i,p} = \mathbf{R}_{i,p}\hat{\mathbf{J}}_{i,p}^H$ and $\mathbf{W}_{i,p} = \mathbf{R}_{i,p}\mathbf{R}_{i,p}^H$. Each sensor gain element is thus estimated by

$$[\hat{\mathbf{g}}_p]_k = \Big(\sum_{i=1}^{D}[\mathbf{W}_{i,p}^*]_{k,k}\Big)^{-1}\sum_{i=1}^{D}[\mathbf{X}_{i,p}^*]_{k,k}. \tag{3.48}$$

- Estimation of φ_i: To do so, we consider the following minimization problem

$$\hat{\varphi}_i = \underset{\varphi_i}{\operatorname{argmin}}\, \tilde{\kappa}(\varphi_i) \tag{3.49}$$

where $\tilde{\kappa}(\varphi_i) = \sum_{p=1}^{M} ||\hat{\mathbf{J}}_{i,p} - \mathbf{G}_p\mathbf{H}_{i,p}\mathbf{Z}_i(\varphi_i)\mathbf{F}_i||_F^2$. Considering $\partial\tilde{\kappa}(\varphi_i)/\partial\varphi_i$ and setting the result to zero leads to

$$\sum_{p=1}^{M} \mathrm{Tr}\Big\{ j\exp^{-j\hat{\varphi}_i}\hat{\mathbf{J}}_{i,p}\mathbf{F}_i^H\mathbf{H}_{i,p}^H\mathbf{G}_p^H - j\exp^{j\hat{\varphi}_i}\mathbf{G}_p\mathbf{H}_{i,p}\mathbf{F}_i\hat{\mathbf{J}}_{i,p}^H \Big\} = 0 \tag{3.50}$$

and finally

$$\exp\Big\{2j\hat{\varphi}_i\Big\} = \frac{\mathrm{Tr}\Big\{\mathbf{M}_{i,p}\Big\}}{\mathrm{Tr}\Big\{\mathbf{M}_{i,p}^H\Big\}} \tag{3.51}$$

with $\mathbf{M}_{i,p} = \sum_{p=1}^{M}\hat{\mathbf{J}}_{i,p}\mathbf{F}_i^H\mathbf{H}_{i,p}^H\mathbf{G}_p^H$ and from which we can directly deduce $\hat{\varphi}_i$.

- Estimation of ϑ_i: We address the following one-dimensional minimization problem

$$\hat{\vartheta}_i = \underset{\vartheta_i}{\operatorname{argmin}} \sum_{p=1}^{M} ||\hat{\mathbf{J}}_{i,p} - \mathbf{G}_p\mathbf{H}_{i,p}\mathbf{Z}_i\mathbf{F}_i(\vartheta_i)||_F^2 \tag{3.52}$$

for each source. Estimates can be obtained in a reasonable computational time through classical data grid search followed by Newton type algorithm.

The global procedure of the RCMLE in the structured case is exposed in Algorithm 4.

Algorithm 4: RCMLE Structured case

input : D, M, B, $\mathbf{C}_i$, β_i, $\mathbf{y}$, $\hat{\mathbf{J}}_{i,p}$ as output of Algorithm 3, $i \in \{1,\dots,D\}$ and $p \in \{1,\dots,M\}$
output : $\hat{\boldsymbol{\varepsilon}}$
initialize: $\hat{\boldsymbol{\varepsilon}} \leftarrow \boldsymbol{\varepsilon}_{\text{init}}$
while *stop criterion unreached* **do**
1 Obtain $\hat{\vartheta}_i$ from (3.52), $i \in \{1,\dots,D\}$
2 Obtain $\hat{\mathbf{g}}_p$ from (3.48), $p \in \{1,\dots,M\}$
3 Obtain $\hat{\varphi}_i$ from (3.51), $i \in \{1,\dots,D\}$

3.5 Simulation results

In this section, we aim to evaluate the performance of the designed algorithm: we start with a CG-based scenario where the noise model matches our noise assumption and we plot the corresponding CRB. Afterwards, we intend to compare, in a more realistic scenario, the proposed algorithm with the recently introduced approach based on the Student's t [2] and the traditional Gaussian cases [14]. We study both non-structured and structured situations and also compare the relaxed deterministic technique with Bayesian alternatives. Finally, visual performance is provided on academic examples by creating residual images where weak background sources are revealed.

3.5.1 Under CG-based noise

Firstly, we aim to assess the estimation performance of the proposed CG-based RCMLE and its numerical stability. We consider the non-structured case where $\boldsymbol{\theta}$ in (3.32) corresponds to the real and imaginary parts of the entries of all Jones matrices. We assume that the additive noise in (3.5) follows a CG distribution as in (3.2), *i.e.*, the generated noise matches the noise model assumption. We choose to generate each random texture component from an inverse Gamma distribution [144], as follows

$$\tau_{pq} \sim \mathcal{IG}(\nu/2, \nu/2), \tag{3.53}$$

with ν degrees of freedom [16] and $[\boldsymbol{\Omega}]_{k,l} = \sigma^2 0.9^{|k-l|} \exp^{j\frac{\pi}{2}(k-l)}$. With (3.53), the resulting noise distribution for $\mathbf{n}_{pq}$ is a Student's t.

To assess the statistical performance, we make use of the CRB [145]. Under quite general/weak conditions, the variance satisfies

$$\text{MSE}([\hat{\boldsymbol{\theta}}]_k) = \text{E}\Big\{ \Big([\hat{\boldsymbol{\theta}}]_k - [\boldsymbol{\theta}]_k \Big)^2 \Big\} \geqslant [\text{CRB}(\boldsymbol{\theta})]_{k,k} \tag{3.54}$$

where the CRB is the inverse of the FIM, noted $\mathbf{F}$. In our specific case, we adapt the Slepian-bangs type formula to our context [146, 147] and obtain

$$[\mathbf{F}]_{k,l} = 2\frac{\nu+4}{\nu+5} \sum_{pq} \Re \left\{ \frac{\partial \tilde{\mathbf{v}}_{pq}^H(\boldsymbol{\theta})}{\partial [\boldsymbol{\theta}]_k} \boldsymbol{\Omega}^{-1} \frac{\partial \tilde{\mathbf{v}}_{pq}(\boldsymbol{\theta})}{\partial [\boldsymbol{\theta}]_l} \right\}. \tag{3.55}$$

Noise parameters being decoupled from parameters of interest $\boldsymbol{\theta}$, we only keep the part of the FIM corresponding to the latter.

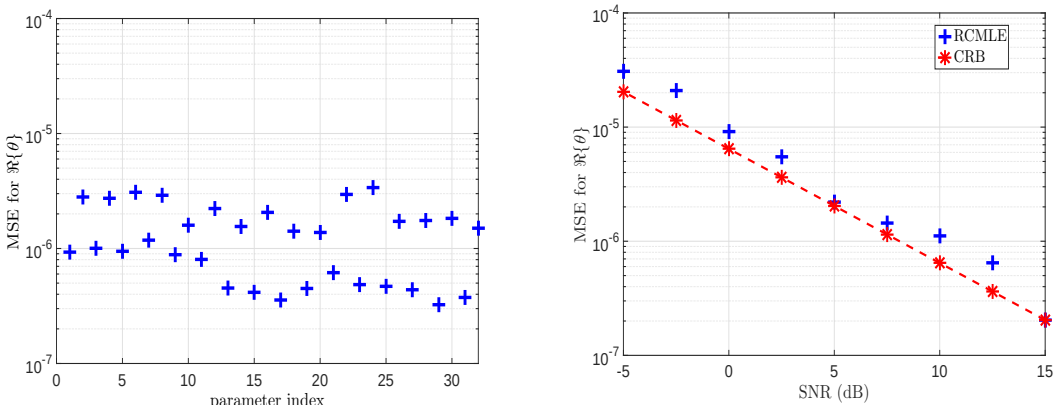
We consider $D = 2$ calibration sources and $M = 8$ sensors in the array, so the total number of real unknowns is given by $8DM = 128$ and the number of measurements is $8B = 224$. Prior information about the source $\mathbf{C}_i$ is generated thanks to random Stokes parameters [5, 18]. In Figure 3.6(a), we plot the MSE of the real part of unknown entries corresponding to the first source, leading to 32 estimated points. The Signal-to-Noise Ratio (SNR) is fixed at 15 dB and the behavior remains the same for the imaginary part or the other source. For the real part of the unknown parameter vector $\boldsymbol{\theta}$, the MSE is defined as follows

$$\text{MSE} = \frac{1}{MC} \sum_{k=1}^{MC} \Re \Big\{ (\hat{\boldsymbol{\theta}}_k - \boldsymbol{\theta})^2 \Big\} \tag{3.56}$$

where MC is the number of Monte-Carlo runs, which we usually fix to 100 in our work. In Figure 3.6(b), we plot the MSE of a given unknown parameter as a function of the SNR, with the corresponding CRB deduced from (3.55). Let us note that the MSE gets close to the CRB but a slight gap remains. This could be explained by our relaxed version of the exact ML estimator. Indeed, the RCMLE derives estimates for unknown and deterministic texture parameters while they were in fact generated as inverse Gamma distributed random parameters (3.53) in the data model and the CRB relies on this specific prior distribution.

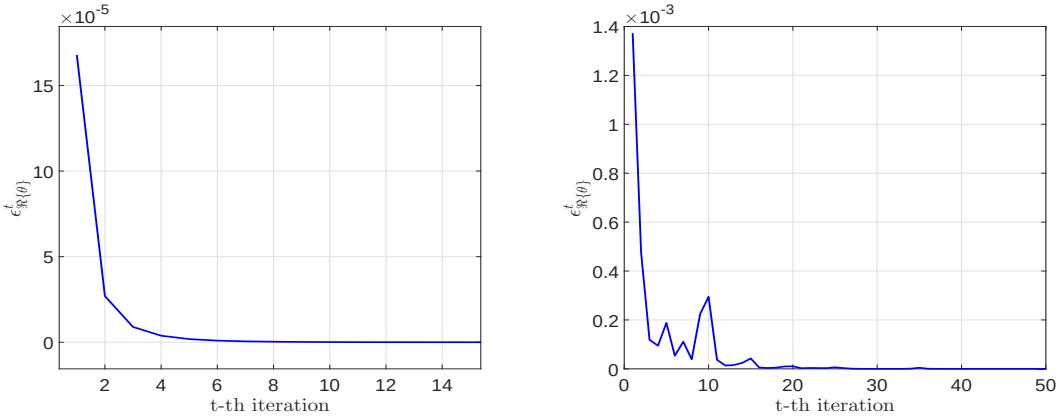
In Figure 3.7, we investigate numerical convergence rates of the RCMLE, for each of its three loops as exposed in Algorithm 3, and see how estimation of $\boldsymbol{\theta}$ is affected through time. To do so, we introduce the following quantity

$$\epsilon^t_{\Re\{\boldsymbol{\theta}\}} = ||\Re\left\{\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t-1}\right\}||_2^2. \tag{3.57}$$



(a) (b)

Figure 3.6: (a) MSE of the real part of the first 32 unknown parameters for a given SNR, (b) MSE vs. SNR for the real part of a given unknown parameter and the corresponding CRB, for $D = 2$ bright signal sources and $M = 8$ antennas.



(a) (b)

Figure 3.7: Numerical stability for the second loop (a), and the first loop (b) in Algorithm 3.

We plot $\epsilon^t_{\Re\{\boldsymbol{\theta}\}}$ as function of the t-th iteration for the second and first loops, in Figures 3.7(a) and 3.7(b), respectively. Let us notice that the third loop is not reported here as it has the same behavior as the convergence in the second loop. From Figure 3.7(a), we deduce that a few iterations are enough to reach stability while from Figure 3.7(b), approximately 20 iterations are required to attain convergence. In practical, only 3 to 4 iterations were needed to approach the CRB.

3.5.2 Under realistic scenario

We aim to study the designed algorithm in a realistic scenario with D calibration sources and D' weak background sources which act as outliers and contribute to the noise component, as shown in (2.47) and (3.1). The SNR is defined as the ratio of the normalized power of D calibration sources over the sum of normalized power of D' background sources and a power noise factor. Jones matrices stand for the perturbation effects along the signal propogation path

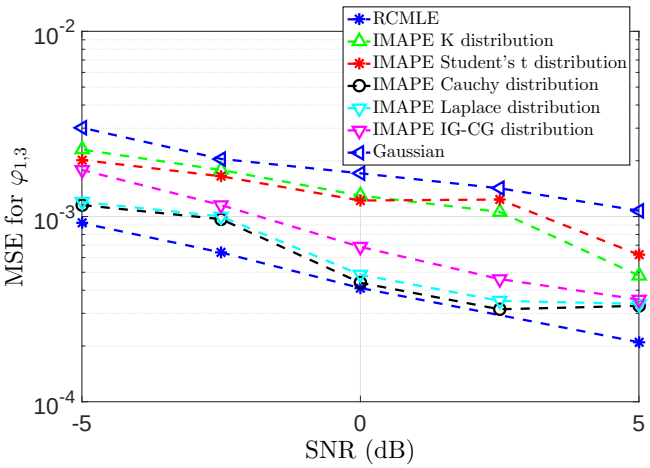


Figure 3.8: MSE for a given ionospheric phase delay as function of the SNR for different kinds of CG distributions.

and include in particular the electronic gains in (2.17), the ionospheric phase shifts in (2.16) and the Faraday rotation in (2.18), which are all randomly generated here.

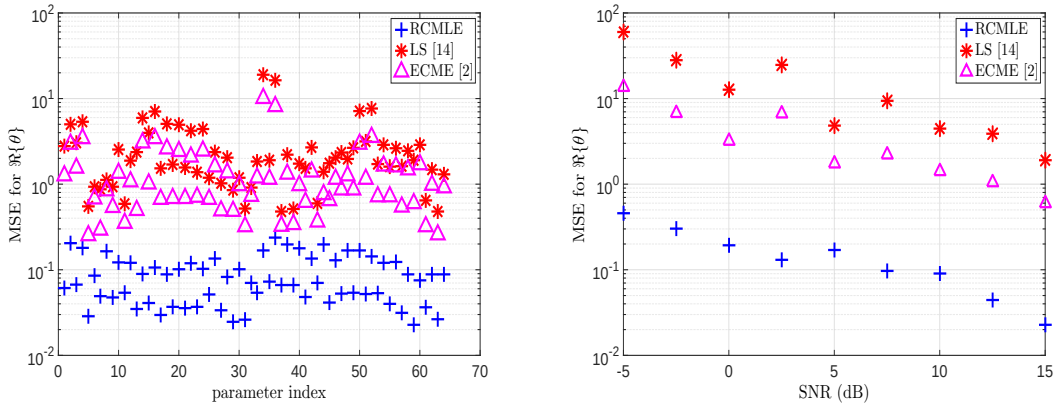
3.5.2.1 Bayesian vs. deterministic

We introduced different kinds of CG-based noise models obtained with different texture priors in section 3.3.1. To compare their estimation performances with the RCMLE and see how estimation of parameters of interest $\boldsymbol{\theta}$ is affected, let us consider a realistic scenario with $D = 2$ bright calibration sources, $M = 8$ antennas in the sensor array and $D' = 4$ weak background sources, resulting in non-Gaussian environment.

In Figure 3.8, we plot the MSE of the previously introduced estimators for one representative ionospheric phase delay, *i.e.*, $\varphi_{1,3}$ here (first source, third antenna), as a function of the SNR. To assess the robustness of the proposed methods, we also plot the MSE for traditional calibration based on a Gaussian noise assumption and the minimization of a LS cost function. As expected, this restricted hypothesis leads to poor performance while estimation results differ depending on which $p(\tau_{pq}; \boldsymbol{\varphi})$ is considered. Still, we notice that the lowest MSE is achieved with the RCMLE which enables to reach robustness due to its flexibility and lack of misspecifications, resulting in more accuracy. Therefore, the relaxed version of the ML estimator, with texture components considered as deterministic in the estimation process, will be favoured in the rest of the study.

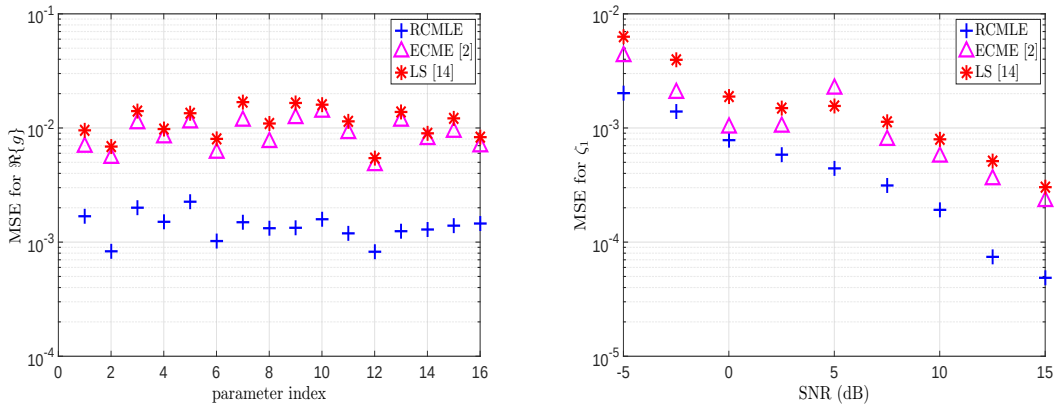
3.5.2.2 State-of-the-art vs. deterministic

To evaluate and compare the performances, we consider the RCMLE from Algorithm 3, the robust approach based on the specific Student's t distribution [2] and the traditional Gaussian noise assumption [14] which both compose state-of-the-art. As a first step, we still study the non-structured case in which $\boldsymbol{\theta}$ includes real and imaginary parts of all Jones matrices along the calibration source paths. The approach in [2] specifically considers a Student's t noise modeling with i.i.d. entries and uses the ECME algorithm as described in section 2.4.3.3 [110, 148]. Traditional calibration is based on the zero-mean white Gaussian noise assumption and solves a LS problem as in section 2.4.2 [14]. Results are plotted in Figure 3.9 for similar computation times. Better accuracy is obtained with the RCMLE as no specific noise distribution has been chosen and CG-based models include a wide range of different distributions, leading to more flexibility



(a) (b)

Figure 3.9: (a) MSE of the real part of the 64 unknown parameters for a given SNR, (b) MSE vs. SNR for the real part of a given unknown parameter, for $D = 2$, $M = 8$ and $D' = 8$, leading to 128 real parameters of interest to estimate and 224 measurements.



(a) (b)

Figure 3.10: (a) MSE of the real part of the 16 complex gains for a given SNR, (b) MSE of ζ_1 vs. SNR, for $D = 2$, $M = 8$, and $D' = 4$, leading to 38 real parameters of interest to estimate and 224 measurements.

and robustness. Besides, no assumption has been made about independent entries within the noise vector $\mathbf{n}_{pq}$, unlike [2]. Let us remark that in Figure 3.9(a), the MSE is plotted for both $D = 2$ calibration sources, leading to $4DM = 64$ plotted points for the real part of $\boldsymbol{\theta}$.

3.5.3 Under structured case

Once all Jones entries in $\boldsymbol{\theta}$ have been estimated, we aim to apply the algorithm for structured Jones matrices, exposed in Algorithm 4. Both algorithms based on the Student's t [2] and the Gaussian distribution [14] have been introduced in the non-structured case so we apply the structured case on the output of these two approaches. Jones matrices are randomly generated according to (3.6) and we introduce $\mathbf{g} = [\mathbf{g}_1^T, \ldots, \mathbf{g}_M^T]^T$. For the ionospheric phase delay in (3.7), it is generated as follows [25]

$$\varphi_{i,p} = \eta_i u_p + \zeta_i v_p. \tag{3.58}$$

Equivalently, we can write

$$\boldsymbol{\varphi}_i^T = \boldsymbol{\alpha}_i^T \boldsymbol{\Lambda} \tag{3.59}$$

where $\boldsymbol{\varphi}_i = [\varphi_{i,1}, \ldots, \varphi_{i,M}]^T$, $\boldsymbol{\Lambda} = \begin{bmatrix} u_1, & \ldots & , u_M \\ v_1, & \ldots & , v_M \end{bmatrix}$, $\boldsymbol{\alpha}_i = [\eta_i, \zeta_i]^T$ refers to the two-dimensional offsets by which the i-th source is shifted and known position $\mathbf{r}_p = [u_p, v_p]^T$ of the p-th antenna is expressed in wavelength units. Once $\varphi_{i,p}$ has been estimated with (3.51) for $p \in \{1, \ldots, M\}$, we can deduce directional shifts $\hat{\boldsymbol{\alpha}}_i$ with the following aditional step

$$\hat{\boldsymbol{\alpha}}_i^T = \frac{\hat{\boldsymbol{\varphi}}_i^T \boldsymbol{\Lambda}^H \begin{bmatrix} \sum_{p=1}^M v_p^2 & -\sum_{p=1}^M u_p v_p \\ -\sum_{p=1}^M v_p u_p & \sum_{p=1}^M u_p^2 \end{bmatrix}}{\sum_{p=1}^M u_p^2 \sum_{p=1}^M v_p^2 - (\sum_{p=1}^M u_p v_p)^2}. \tag{3.60}$$

The MSE results are shown in Figure 3.10, with similar computation times, for specific physical parameters: the complex gains for fixed SNR and the source offset ζ_1 as function of the SNR, the behavior being the same for other parameters. Better performances are obtained with the proposed RCMLE in the structured case. This was already expected from Figure 3.9 as better estimation of Jones entries naturally leads to better estimation of parameters describing them.

3.5.4 Recovered images

In this section, we investigate the formation of residual images on academic examples, obtained with *Matlab*, but more realistic data simulation and image recovering will be presented in Chapter 5.

Let us consider $D = 2$ calibration sources in the sky model. In the ideal case, visibilities are only affected by the **K**-Jones term in (2.15). But due to perturbation effects, especially the ionospheric phase delays in (3.58), source positions are subject to shifts and the apparent location may differ from the exact one. We wish to recover $D' = 4$ background sources whose exact intensities and positions are unknown to us and randomly generated from a discrete uniform distribution. Only ionospheric phase delays are incorporated within the Jones chain (2.13).

Calibration is performed with three different methods: the proposed RCMLE, the Student's t [2] and the Gaussian cases [14]. From the phase estimates, we deduce the offsets with (3.60) by which the $D = 2$ calibration sources are shifted. Then, the corresponding offsets for weak sources are deduced by linear interpolation. More specifically, we are interested in residual visibilities.

Indeed, from (2.47), for each antenna pair, we know that

$$\mathbf{v}_{pq} = \sum_{i=1}^{D} \mathbf{s}_{i,pq}(\boldsymbol{\theta}) + \sum_{i'=1}^{D'} \mathbf{s}_{i',pq} + \mathbf{n}_{pq} \tag{3.61}$$

with the calibration, the non-calibration and the noise components, respectively. By residual visibilities, we mean the following quantity

$$\mathbf{v}_{pq}^{RES} = \mathbf{v}_{pq} - \sum_{i=1}^{D} \mathbf{s}_{i,pq}(\hat{\boldsymbol{\theta}}) \tag{3.62}$$

in which $\boldsymbol{\theta}$ is estimated using one of the three aformentionned estimators. The true situation corresponds to the case when $\boldsymbol{\theta}$ is exactly known. Intensities of weak sources are then recovered by performing a two-dimensional Inverse Fourier Transform (2.6) on the residual visibilities.

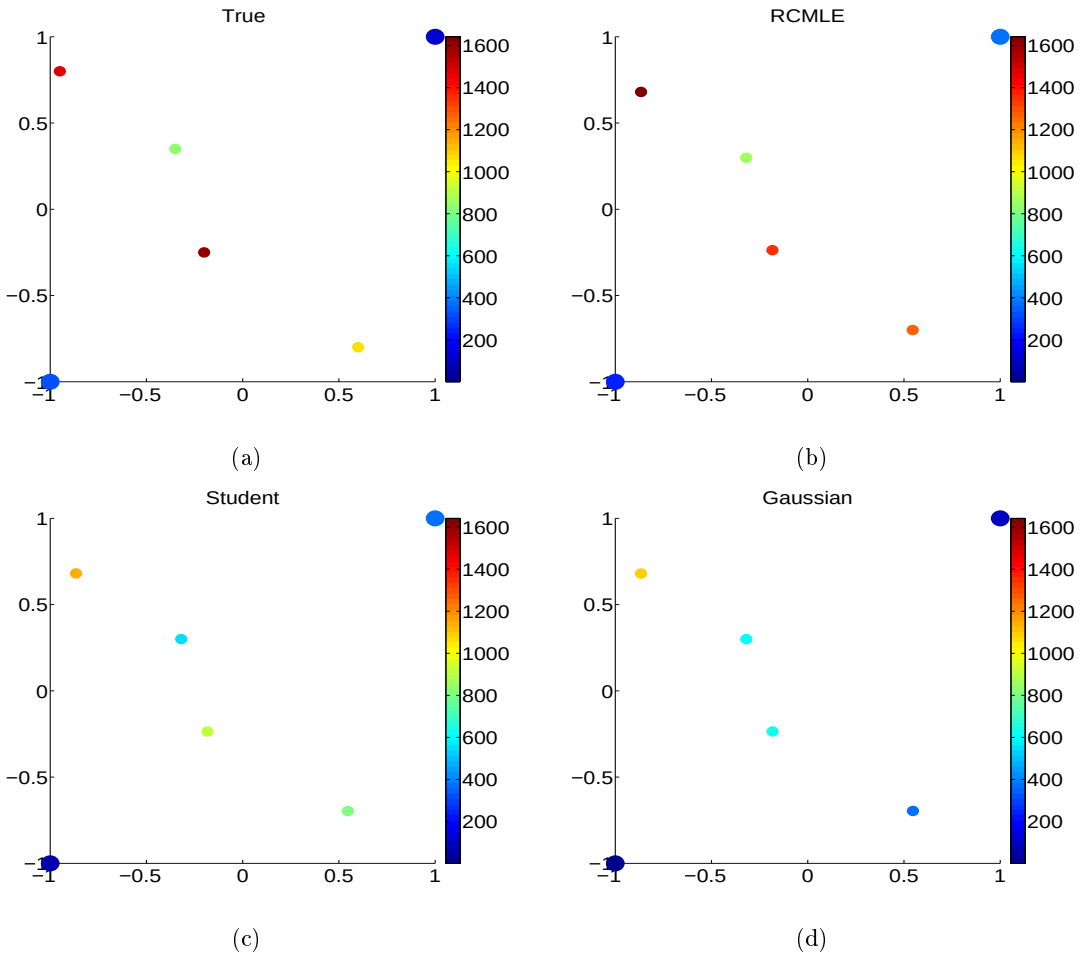


Figure 3.11: Recovered residual images when Jones matrices are (a) perfectly estimated (b) estimated with the RCMLE, (c) estimated with the Student's t noise model and (d) the Gaussian noise assumption (point-like sources in sky image).

Table 3.2: Mean difference between true and estimated positions for the D' background sources.

	x	y
RCMLE	0.0030	0.0709
Student's t case	0.0057	0.0717
Gaussian case	0.0066	0.0721

Recovered intensities at the calibration and non-calibration positions are represented thanks to colorbars in Figure 3.11 for a mostly empty sky with point-like sources. Thus, if calibration is performed exactly, there should be no flux left at the calibration positions. Position shifts being hardly visible, to make it more clear, we compute in Table 3.2 the mean over all D' sources of the difference between the exact source location and the estimated one, in both directions (x, y). The lower is this difference, the more accurate is the calibration method. We recover similar residual images in Figure 3.12 but intensity is shown for each position of the plane.

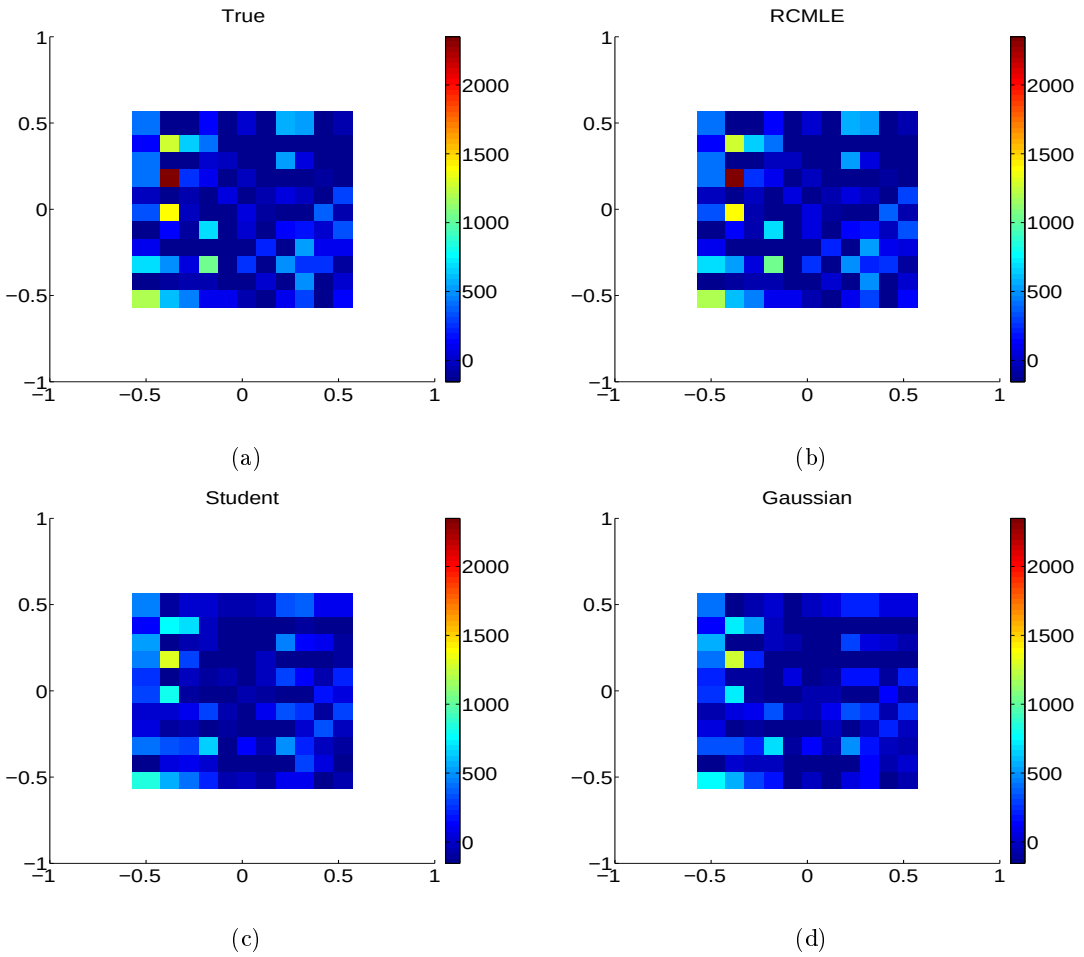


Figure 3.12: Recovered residual images when Jones matrices are (a) perfectly estimated (b) estimated with the RCMLE, (c) estimated with the Student's t noise model and (d) the Gaussian noise assumption.

3.6 Conclusion

In radio astronomy, Jones matrices model all the perturbation effects which corrupt the observations and they can be considered as structured or not depending on the scenario under study. However, visibility measurements are also affected by the presence of outliers wich change the typical Gaussian nature of noise. To achieve robustness, we propose the RCMLE based on the CG model and a relaxed assumption. This algorithm reveals to be more robust for both non-structured and structured cases, it is less subject to errors than MAP approaches and benefits from reasonable computational complexity. Indeed, using the EM and the BCD algorithms enables to reduce the global computational cost thanks to a per source and per antenna parameter partition, leading to multiple sub-problems instead of a global multi-dimensional one. Texture, speckle components and Jones matrices are estimated iteratively thanks to closed-form expressions. However, their specific structure of variation, notably *w.r.t.* frequency, is not taken into account and could be exploited to enhance the calibration process. Thus, the extension to the multi-frequency scenario is studied in the following chapter.

Chapter 4

Robust distributed calibration in multi-frequency scenario

4.1 Motivation for multi-frequency calibration

The new generation of radio interferometers requires computationally efficient algorithms wich perform calibration accurately while being robust. Robustness was addressed in Chapter 3 by proposing the CG model for the noise contribution. Unknown parameters of interest were estimated for one given frequency but in practice, radio interferometers operate on a broad spectral range, resulting in large multi-frequency data from which information is retrieved. Perturbation effects, modeled by Jones matrices, as well as the noise contribution do not remain constant across frequency and their variation can be exploited during calibration in a multi-frequency scenario [149]. In this chapter, we focus on the frequency dependence (in addition, one might take into account the temporal variation using the same approach by considering independent solutions between time intervals). We extend the structured case, whose model was exposed in section 3.2.2 and the corresponding estimator in section 3.4.2, to regime 4 which is more adapted to large interferometric arrays. Thus, propagation conditions are variable per station and per source due to large field-of-view and long baselines, leading to more unknowns to estimate than in regime 3 where the same part of the ionosphere is seen by all antennas in the station. Let us note that adaptation to regime 3 is straightforwardly obtained and multi-frequency calibration for the non-structured case is discussed in [35].

Considering multiple frequency channels can significantly increase the generated data volumes so the flow of information needs to be handled in the most optimal way. To this end, distributed calibration can be exploited with a network of computational agents. Each of them achieves calibration locally for a given subband and a fusion center collects data from all different frequencies. This method benefits from data parallelism and distributed computing load, contrary to standard arrangement where a single computational agent directly operates on the full observing bandwidth and considers one frequency bin at a time [4, 12]. Furthermore, making use of distributed procedures is consistent with the new radio synthesis arrays since the global bandwidth is divided into multiple frequency channels. Exploiting frequency diversity is not new and has already been tackled with bandpass solutions [150] or smooth polynomials [35] in the non-structured scenario. But the purpose of this chapter is to combine robustness with estimation of physical parameters taking into account their structured variation across frequency in regime 4.

4.2 Variation of parameters across frequency

We are now dealing with interferometer calibration where direction independent and direction dependent effects need to be corrected. The Jones chain given in (3.6) for regime 3 is adapted to regime 4 in a multi-frequency scenario as follows

$$\mathbf{J}_{i,p}^{[f]}(\boldsymbol{\theta}^{[f]}) = \mathbf{G}_{p}^{[f]}(\mathbf{g}_{p}^{[f]})\mathbf{H}_{i,p}^{[f]}\mathbf{Z}_{i,p}^{[f]}(\varphi_{i,p}^{[f]})\mathbf{F}_{i,p}^{[f]}(\vartheta_{i,p}^{[f]}) \tag{4.1}$$

in which the frequency dependence of ionospheric distortions can be accurately described. Indeed, the ionospheric phase delay in (3.7) is specified now as [33]

$$\varphi_{i,p}^{[f]} \propto \frac{TEC_{i,p}}{f} \tag{4.2}$$

where TEC was defined in section 2.2.1. As regards the Faraday rotation in (3.8), it is given by [25]

$$\vartheta_{i,p}^{[f]} \propto \frac{RM_{i,p}}{f^2} \tag{4.3}$$

in which RM refers to the rotation measure, depending on the magnetic field and the density of electrons along the propagation path i-p.

There is no available model for the frequency variation of the gains $\mathbf{g}_{p}^{[f]}$. However, we can enforce smoothness across frequency by introducing polynomial variation across frequency, as discussed in section 4.3.2.

Therefore, for one given frequency $f \in \mathcal{F} = \{f_1, \ldots, f_F\}$, the number of unknown parameters of interest is gathered into the $(2MD + 2M) \times 1$ vector $\boldsymbol{\theta}^{[f]} = [\boldsymbol{\epsilon}^{[f]^T}, \mathbf{g}_1^{[f]^T}, \ldots, \mathbf{g}_M^{[f]^T}]^T$ where $\boldsymbol{\epsilon}^{[f]} = [\vartheta_{1,1}^{[f]}, \ldots, \vartheta_{D,M}^{[f]}, \exp(j\varphi_{1,1}^{[f]}), \ldots, \exp(j\varphi_{D,M}^{[f]})]^T$ collects frequency dependent per-receiver and per-source ionospheric effects. Furthermore, we still consider the noise model in (3.2) so we also need to estimate the $B \times 1$ texture realizations $\boldsymbol{\tau}^{[f]} = [\tau_{12}^{[f]}, \tau_{13}^{[f]}, \ldots, \tau_{(M-1)M}^{[f]}]^T$ and the 4×4 speckle covariance matrix $\boldsymbol{\Omega}^{[f]}$. We notice that frequency dependence has been added for the noise parameters as well but we do not assume any specific prior structure *w.r.t.* f. Nevertheless, if a priori knowledge about noise variation *w.r.t.* frequency is available, it can be straightforwardly incorporated into the proposed algorithm.

4.3 Estimation of Jones matrices in structured case

4.3.1 Robust estimation in multi-frequency scenario

Assuming independence between $\mathbf{n}_{pq}^{[f]}$, the corresponding expression of the log-likelihood in (3.12) reads

$$\ln p\left(\{\mathbf{y}^{[f]}\}_{f\in\mathcal{F}} \mid \{\boldsymbol{\theta}^{[f]}, \boldsymbol{\tau}^{[f]}, \boldsymbol{\Omega}^{[f]}\}_{f\in\mathcal{F}}\right) = -\sum_{f\in\mathcal{F}} l^{[f]}(\boldsymbol{\theta}^{[f]}) - \sum_{f\in\mathcal{F}}\sum_{pq} \ln|\pi\tau_{pq}^{[f]}\boldsymbol{\Omega}^{[f]}| \tag{4.4}$$

with

$$l^{[f]}(\boldsymbol{\theta}^{[f]}) = \sum_{pq} \frac{1}{\tau_{pq}^{[f]}} \mathbf{u}_{pq}^{[f]^H}(\boldsymbol{\theta}^{[f]})\boldsymbol{\Omega}^{[f]^{-1}}\mathbf{u}_{pq}^{[f]}(\boldsymbol{\theta}^{[f]}). \tag{4.5}$$

Then, we adopt the same procedure as for the RCMLE and obtain similar texture and speckle covariance estimates than in (3.27) and (3.29). More specifically, we can write

$$\hat{\tau}_{pq}^{[f]} = \frac{1}{4}\mathbf{u}_{pq}^{[f]^H}(\boldsymbol{\theta}^{[f]})\boldsymbol{\Omega}^{[f]^{-1}}\mathbf{u}_{pq}^{[f]}(\boldsymbol{\theta}^{[f]}) \tag{4.6}$$

for the texture, and

$$\hat{\mathbf{\Omega}}^{[f]} = \frac{4}{B} \sum_{pq} \frac{\mathbf{u}_{pq}^{[f]}(\boldsymbol{\theta}^{[f]}) \mathbf{u}_{pq}^{[f]^H}(\boldsymbol{\theta}^{[f]})}{\mathbf{u}_{pq}^{[f]^H}(\boldsymbol{\theta}^{[f]})(\hat{\mathbf{\Omega}}^{[f]})^{-1} \mathbf{u}_{pq}^{[f]}(\boldsymbol{\theta}^{[f]})} \tag{4.7}$$

for the speckle, followed by the normalization step $\hat{\mathbf{\Omega}}^{[f]} \leftarrow \frac{\hat{\mathbf{\Omega}}^{[f]}}{\mathrm{tr}\left\{\hat{\mathbf{\Omega}}^{[f]}\right\}}$ to avoid any ambiguity. The principle of the proposed multi-frequency robust approach is similar to the one exposed in Algorithm 2: we estimate alternatively $\{\hat{\boldsymbol{\theta}}^{[f]}\}_{f\in\mathcal{F}}$, $\{\hat{\mathbf{\Omega}}^{[f]}\}_{f\in\mathcal{F}}$ and $\{\hat{\boldsymbol{\tau}}^{[f]}\}_{f\in\mathcal{F}}$ in a global iterative loop. In what follows, we focus on estimation of $\{\boldsymbol{\theta}^{[f]}\}_{f\in\mathcal{F}}$, for a fixed $\{\hat{\boldsymbol{\tau}}^{[f]}\}_{f\in\mathcal{F}}$ given by (4.6) and $\{\hat{\mathbf{\Omega}}^{[f]}\}_{f\in\mathcal{F}}$ given by (4.7), by considering consensus-based distributed optimization [34, 36].

4.3.2 Distributed estimation in multi-frequency scenario

Multi-frequency calibration considers a whole frequency range with multiple sub-frequency bands. A computationally efficient way to handle this situation is to apply decentralized strategies with distributed and consensus algorithms as the ADMM which was introduced in section 2.4.3.4. This technique is well-suited for large-scale problems and has already been applied for image reconstruction [117]. We will be considering this procedure in the following with a group of computational agents. We assume that each of them has access to data for a specific frequency $f \in \mathcal{F} = \{f_1, \ldots, f_F\}$ and solves a local problem. Estimates are then transmitted to the fusion center which enforces consensus among agents and finally transfers the updates to each local agent. Compared to independent (per-channel) calibration performed for one frequency after another, distributed calibration can lead to better accuracy since information across all frequencies is taken into account.

The idea is to be able to get closed-form expressions for the parameters of interest. To this end, for $i \in \{1, \ldots, D\}$ and $p \in \{1, \ldots, M\}$, we rewrite the ionospheric phase delay as

$$\exp\left(j\varphi_{i,p}^{[f]}\right) = \mathbf{b}^{[f]^T} \mathbf{z}_{i,p} \tag{4.8}$$

and the Faraday rotation angle as

$$\vartheta_{i,p}^{[f]} = \frac{1}{f^2} \overline{z}_{i,p} \tag{4.9}$$

s.t. the $(M + NM) \times 1$ unknown frequency independent vector of hidden variables $\mathbf{z}_i = [\overline{z}_{i,1}, \ldots, \overline{z}_{i,M}, \mathbf{z}_{i,1}^T, \ldots, \mathbf{z}_{i,M}^T]^T$ is introduced for the i-th source direction. The frequency dependent vector $\mathbf{b}^{[f]}$ is defined *s.t.*

$$[\mathbf{b}^{[f]}]_k = \left(\frac{1}{f}\right)^{k-1} \tag{4.10}$$

for $k \in \{1, \ldots, N\}$ and $(N-1)$ is the order of the truncated approximation power series of $\exp\left(j\varphi_{i,p}^{[f]}\right)$, chosen arbitrarily. More specifically, we can write

$$\exp\left(j\varphi_{i,p}^{[f]}\right) \simeq \sum_{k=1}^{N} \left(\frac{1}{f}\right)^{k-1} \frac{(j\varphi_{i,p}^{[f_0]})^{k-1}}{(k-1)!} \tag{4.11}$$

s.t. $\varphi_{i,p}^{[f]} = \frac{\varphi_{i,p}^{[f_0]}}{f}$ and $\varphi_{i,p}^{[f_0]} \propto TEC_{i,p}$. Let us note that $\overline{z}_{i,p} \propto RM_{i,p}$.

As mentioned before, we impose smoothness over frequency for the gains, *s.t.*,

$$\mathbf{g}_p^{[f]} = \tilde{\mathbf{B}}^{[f]} \tilde{\mathbf{z}}_p \tag{4.12}$$

where the $2 \times 2\tilde{N}$ frequency modeling matrix reads

$$\tilde{\mathbf{B}}^{[f]} = \tilde{\mathbf{b}}^{[f]^T} \otimes \mathbf{I}_2 \tag{4.13}$$

and the $\tilde{N} \times 1$ vector $\tilde{\mathbf{b}}^{[f]}$ is described by

$$[\tilde{\mathbf{b}}^{[f]}]_k = \left(\frac{f - f_0}{f_0}\right)^{k-1} \tag{4.14}$$

with $k \in \{1, \ldots, \tilde{N}\}$, an arbitrarily chosen $\tilde{N}$ and a given reference frequency f_0 [35]. Frequency independent variables are introduced as well for the p-th receiving element through the $2\tilde{N} \times 1$ vector $\tilde{\mathbf{z}}_p$.

With (4.4), (4.8), (4.9) and (4.12), the calibration problem can be formulated as

$$\{\hat{\boldsymbol{\theta}}^{[f]}\}_{f\in\mathcal{F}}, \hat{\mathbf{z}} = \underset{\{\boldsymbol{\theta}^{[f]}\}_{f\in\mathcal{F}}, \mathbf{z}}{\operatorname{argmin}} \sum_{f\in\mathcal{F}} l^{[f]}(\boldsymbol{\theta}^{[f]}) \quad s.t. \quad \boldsymbol{\theta}^{[f]} = \mathbf{B}^{[f]}\mathbf{z} \tag{4.15}$$

with $\mathbf{z} = [\overline{z}_{1,1}, \ldots, \overline{z}_{D,M}, \mathbf{z}_{1,1}^T \ldots, \mathbf{z}_{D,M}^T, \tilde{\mathbf{z}}_1^T, \ldots, \tilde{\mathbf{z}}_M^T]^T$ the $(DM + DNM + 2\tilde{N}M) \times 1$ augmented vector of hidden variables and the frequency coupling matrix is given by

$$\mathbf{B}^{[f]} = \text{bdiag}\Big\{\frac{1}{f^2}\mathbf{I}_{DM}, \mathbf{I}_{DM} \otimes \mathbf{b}^{[f]^T}, \mathbf{I}_M \otimes \tilde{\mathbf{B}}^{[f]}\Big\} = \begin{bmatrix} \frac{1}{f^2}\mathbf{I}_{DM} & \mathbf{0}_{DM\times DNM} & \mathbf{0}_{DM\times 2\tilde{N}M} \\ \mathbf{0}_{DM\times DM} & \mathbf{I}_{DM} \otimes \mathbf{b}^{[f]^T} & \mathbf{0}_{DM\times 2\tilde{N}M} \\ \mathbf{0}_{2M\times DM} & \mathbf{0}_{2M\times DNM} & \mathbf{I}_M \otimes \tilde{\mathbf{B}}^{[f]} \end{bmatrix}. \tag{4.16}$$

Therefore, multi-frequency calibration amounts to solve a global constrained optimization problem. To achieve this goal, we introduce the following Lagrangian

$$L\left(\{\boldsymbol{\theta}^{[f]}\}_{f\in\mathcal{F}}, \mathbf{z}, \{\mathbf{x}^{[f]}\}_{f\in\mathcal{F}}\right) = \sum_{f\in\mathcal{F}} L^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right) \tag{4.17}$$

with $L^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right) = l^{[f]}\left(\boldsymbol{\theta}^{[f]}\right) + h^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right)$ where the penalty term reads

$$h^{[f]}\left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]}\right) = 2\Re\left\{\mathbf{x}^{[f]^H}\left(\boldsymbol{\theta}^{[f]} - \mathbf{B}^{[f]}\mathbf{z}\right)\right\} + \rho||\boldsymbol{\theta}^{[f]} - \mathbf{B}^{[f]}\mathbf{z}||_2^2 \tag{4.18}$$

The associated Lagrange parameters (or dual variables) for a given frequency f are denoted as the $(2DM + 2M) \times 1$ vector $\mathbf{x}^{[f]} = [\overline{x}_{1,1}^{[f]}, \ldots, \overline{x}_{D,M}^{[f]}, x_{1,1}^{[f]}, \ldots, x_{D,M}^{[f]}, \tilde{\mathbf{x}}_1^{[f]^T}, \ldots, \tilde{\mathbf{x}}_M^{[f]^T}]^T$ and ρ is a regularization factor.

To sum it up, the ADMM algorithm consists in updating the three following quantities

- $\left(\hat{\boldsymbol{\theta}}^{[f]}\right)^{t+1} = \underset{\boldsymbol{\theta}^{[f]}}{\operatorname{argmin}} L^{[f]}\left(\boldsymbol{\theta}^{[f]}, (\hat{\mathbf{z}})^t, \left(\hat{\mathbf{x}}^{[f]}\right)^t\right)$ performed locally by each agent (4.19)
- $(\hat{\mathbf{z}})^{t+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \sum_{f\in\mathcal{F}} h^{[f]}\left(\left(\hat{\boldsymbol{\theta}}^{[f]}\right)^{t+1}, \mathbf{z}, \left(\hat{\mathbf{x}}^{[f]}\right)^t\right)$ performed globally at the fusion center (4.20)
- $\left(\hat{\mathbf{x}}^{[f]}\right)^{t+1} = \left(\hat{\mathbf{x}}^{[f]}\right)^t + \rho\left(\left(\hat{\boldsymbol{\theta}}^{[f]}\right)^{t+1} - \mathbf{B}^{[f]}(\hat{\mathbf{z}})^{t+1}\right)$ performed locally by each agent (4.21)

where t is the iteration counter. Minimization in (4.20) requires access to local solutions from all agents at every frequency and admits the following closed-form expression

$$\hat{\mathbf{z}} = \left(\rho \sum_{f \in \mathcal{F}} \left(\mathbf{B}^{[f]^H} \mathbf{B}^{[f]} \right) \right)^{-1} \left(\sum_{f \in \mathcal{F}} \mathbf{B}^{[f]^T} \left(\mathbf{x}^{[f]} + \rho \boldsymbol{\theta}^{[f]} \right) \right) \tag{4.22}$$

which is then broadcasted to each agent as a common global variable in order to perform the update in (4.21) locally.

The step of interest is (4.19) and we will be using an iterative approach to estimate $\boldsymbol{\theta}^{[f]}$. On the one hand, we consider $\boldsymbol{\epsilon}^{[f]}$ and minimize successively *w.r.t.* each $\vartheta_{i,p}^{[f]}$ and $\varphi_{i,p}^{[f]}$ for $i \in \{1, \ldots, D\}$ and $p \in \{1, \ldots, M\}$. On the other hand, we will be considering $\mathbf{g}^{[f]}$. Specifically,

• Faraday rotation: Estimation of $\vartheta_{i,p}^{[f]}$ amounts to solve

$$\hat{\vartheta}_{i,p}^{[f]} = \underset{\vartheta_{i,p}^{[f]}}{\operatorname{argmin}} \, L^{[f]} \left(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]} \right) \tag{4.23}$$

which can be computed with a one-dimensional Newton or gradient descent-type algorithm [29]. Parallelization is also possible for all $i \in \{1, \ldots, D\}$ and $p \in \{1, \ldots, M\}$.

• Ionospheric phase delay: For a given $\hat{\vartheta}_{i,p}^{[f]}$ for $i \in \{1, \ldots, D\}$ and $p \in \{1, \ldots, M\}$, minimization of (4.19) *w.r.t.* $\varphi_{i,p}^{[f]}$ results, after some calculus, in

$$\hat{\varphi}_{i,p}^{[f]} = \frac{1}{2} \arg \left\{ -\frac{\alpha_{i,p}^{[f]}}{\beta_{i,p}^{[f]}} \right\}. \tag{4.24}$$

Notations and detailed calculations are provided in Appendix C.

• Electronic gains: In what follows, we omit dependence *w.r.t.* parameters of interest for sake of clarity. Let us recall that $\mathbf{g}_p^{[f]} = \left[[\mathbf{g}_p^{[f]}]_1, [\mathbf{g}_p^{[f]}]_2 \right]^T$, *s.t.*,

$$\mathbf{G}_p^{[f]}(\mathbf{g}_p^{[f]}) = \operatorname{diag}\{\mathbf{g}_p^{[f]}\} = \begin{bmatrix} [\mathbf{g}_p^{[f]}]_1 & 0 \\ 0 & [\mathbf{g}_p^{[f]}]_2 \end{bmatrix}. \tag{4.25}$$

We consider the following derivative

$$\frac{\partial l^{[f]}(\mathbf{g}^{[f]})}{\partial [\mathbf{g}_p^{[f]}]_1} = \sum_{\substack{q=1 \\ q>p}}^{M} \frac{1}{\tau_{pq}^{[f]}} \lambda_{pq}^{[f]} + \sum_{\substack{q=1 \\ q<p}}^{M} \frac{1}{\tau_{qp}^{[f]}} \tilde{\lambda}_{qp}^{[f]} \tag{4.26}$$

where $\mathbf{g}^{[f]} = [\mathbf{g}_1^{[f]^T}, \ldots, \mathbf{g}_M^{[f]^T}]^T$. Calculus and notations are specified in Appendix D, leading to

$$[\hat{\mathbf{g}}_p^{[f]}]_1 = \frac{a_p^{[f]}}{b_p^{[f]}} \tag{4.27}$$

in which

$$a_p^{[f]} = -t_p^{[f]^*} - [\mathbf{g}_p^{[f]}]_2 \left(\sum_{k=1}^{2S} [\mathbf{w}_p^{[f]}]_{2k} [\boldsymbol{\omega}_p^{[f]^*}]_{2k} + \sum_{k=1}^{2V} [\boldsymbol{\varsigma}_p^{[f]}]_{2k} [\boldsymbol{\varrho}_p^{[f]^*}]_{2k} \right) - [\tilde{\mathbf{x}}_p^{[f]}]_1 + \rho \tilde{\mathbf{z}}_p^T [\tilde{\mathbf{B}}^{[f]^T}]_{:,1} \tag{4.28}$$

and

$$b_p^{[f]} = \sum_{k=0}^{2S-1} [\mathbf{w}_p^{[f]}]_{2k+1} [\boldsymbol{\omega}_p^{[f]^*}]_{2k+1} + \sum_{k=0}^{2V-1} [\boldsymbol{\varsigma}_p^{[f]}]_{2k+1} [\boldsymbol{\varrho}_p^{[f]^*}]_{2k+1} + \rho, \tag{4.29}$$

in which we used the *Matlab* notation $[\cdot]_{:,k}$ to refer to the k-th column.

Similarly, estimation of $[\mathbf{g}_p^{[f]}]_2$ is given by

$$[\mathbf{g}_p^{[f]}]_2 = \frac{\tilde{a}_p^{[f]}}{\tilde{b}_p^{[f]}} \tag{4.30}$$

where

$$\tilde{a}_p^{[f]} = -t_p^{[f]^*} - [\mathbf{g}_p^{[f]}]_1 \left(\sum_{k=0}^{2S-1} [\mathbf{w}_p^{[f]}]_{2k+1} [\boldsymbol{\omega}_p^{[f]^*}]_{2k+1} + \sum_{k=0}^{2V-1} [\boldsymbol{\varsigma}_p^{[f]}]_{2k+1} [\boldsymbol{\varrho}_p^{[f]^*}]_{2k+1} \right) - [\tilde{\mathbf{x}}_p^{[f]}]_2 + \rho \tilde{\mathbf{z}}_p^T [\tilde{\mathbf{B}}^{[f]^T}]_{:,2} \tag{4.31}$$

and

$$\tilde{b}_p^{[f]} = \sum_{k=1}^{2S} [\mathbf{w}_p^{[f]}]_{2k} [\boldsymbol{\omega}_p^{[f]^*}]_{2k} + \sum_{k=1}^{2V} [\boldsymbol{\varsigma}_p^{[f]}]_{2k} [\boldsymbol{\varrho}_p^{[f]^*}]_{2k} + \rho. \tag{4.32}$$

The proposed algorithm is based on the CG model in (3.2) and the ADMM procedure for a multi-frequency scenario. We refer to it as the Multi-frequency Robust Calibration Algorithm (MRCA) [151] and the global scheme is exposed in Algorithm 5.

Algorithm 5: MRCA Multi-frequency Robust Calibration Algorithm

initialize: $\{\hat{\boldsymbol{\theta}}^{[f]} \leftarrow \boldsymbol{\theta}_{\text{init}}^{[f]}\}_{f\in\mathcal{F}}$, $\hat{\mathbf{z}} \leftarrow \mathbf{z}_{\text{init}}$, $\{\hat{\mathbf{x}}^{[f]} \leftarrow \mathbf{x}_{\text{init}}^{[f]}\}_{f\in\mathcal{F}}$,
$\{\hat{\boldsymbol{\Omega}}^{[f]} \leftarrow \boldsymbol{\Omega}_{\text{init}}^{[f]}\}_{f\in\mathcal{F}}$,$\{\hat{\boldsymbol{\tau}}^{[f]} \leftarrow \boldsymbol{\tau}_{\text{init}}^{[f]}\}_{f\in\mathcal{F}}$

while *stop criterion unreached* **do**

  while *stop criterion unreached* **do**

    while *stop criterion unreached* **do**

1 Obtain $\{\hat{\vartheta}_{i,p}^{[f]}\}_{i=1,\dots,D,\ p=1,\dots,M}$ locally from (4.23)

2 Obtain $\{\hat{\varphi}_{i,p}^{[f]}\}_{i=1,\dots,D,\ p=1,\dots,M}$ locally from (4.24)

3 Obtain $\hat{\mathbf{g}}^{[f]}$ locally from (4.27) and (4.30)

4 Obtain $\hat{\mathbf{z}}$ globally from (4.22)

5 Obtain $\hat{\mathbf{x}}^{[f]}$ locally from (4.21)

6 Obtain $\{\hat{\boldsymbol{\Omega}}^{[f]}\}_{f\in\mathcal{F}}$ from (4.7)

7 Obtain $\{\hat{\boldsymbol{\tau}}^{[f]}\}_{f\in\mathcal{F}}$ from (4.6)

Remark: Let us note that after step 4 of MRCA in Algorithm 5, it is possible to refine the estimation: from $\hat{\mathbf{z}}$, we deduce an estimation of the $N \times 1$ vector $\mathbf{z}_{i,p}$, associated to the ionospheric phase delay along the path i-p. With (4.8) and (4.11), we derive N estimates for $\varphi_{i,p}^{[f_0]}$ and perform an average. Finally, a new estimate for $\mathbf{z}$ can be deduced and this additional step can be directly incorporated into the estimation procedure.

In Figure 4.1, we represent the operation flow and signaling exchange between the fusion center and each local agent in our estimation procedure.

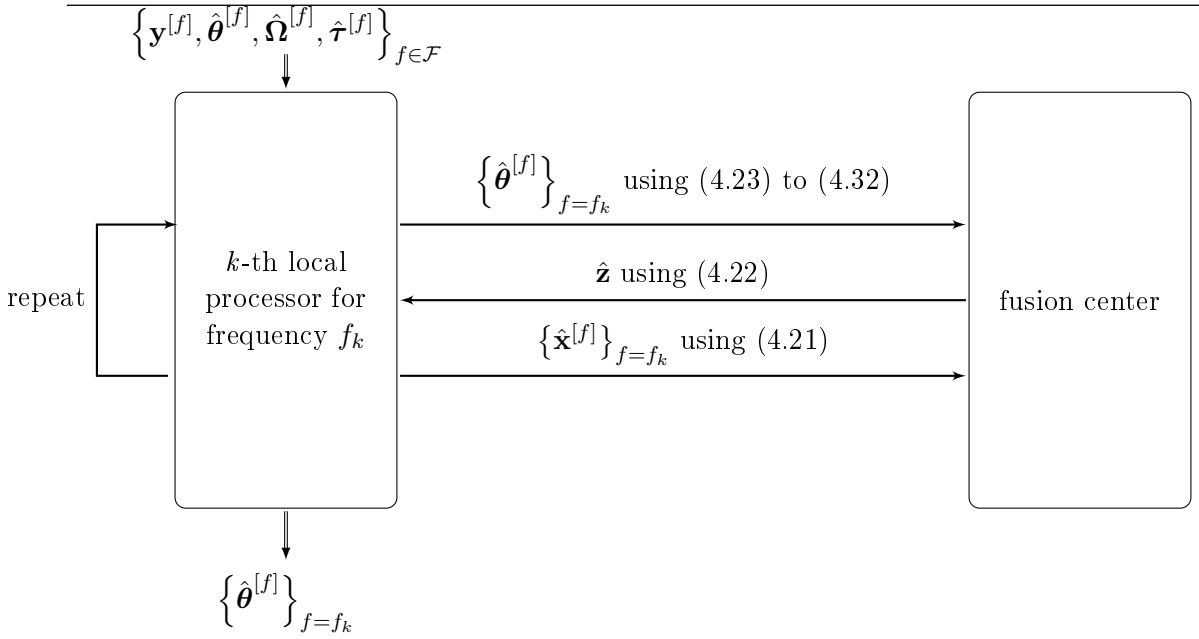


Figure 4.1: Communication between the k-th local processor and the fusion center. The three arrows in the center are performed sequentially and iteratively, which corresponds to the second loop in Algorithm 5.

Let us note that the RCMLE for the structured case exposed in Algorithm 4 was suitable for regime 3 and can be extended to a multi-frequency scenario using the ADMM algorithm as well. Let us call it the Multi-frequency Robust Calibration Algorithm for Regime 3 (MRCA^{R3}). Prior information on $\hat{\mathbf{J}}_{i,p}^{[f]}$ for $i \in \{1,\dots,D\}$, $p \in \{1,\dots,M\}$ and $f \in \mathcal{F}$ is provided by the output of Algorithm 3 and contrary to the original objective function in (4.5), the considered data fidelity term is the following

$$l^{[f]}(\boldsymbol{\theta}^{[f]}) = \sum_{i=1}^{D}\sum_{p=1}^{M} ||\hat{\mathbf{J}}_{i,p}^{[f]} - \mathbf{G}_p^{[f]}(\mathbf{g}_p^{[f]})\mathbf{H}_{i,p}^{[f]}\mathbf{Z}_i^{[f]}(\varphi_i^{[f]})\mathbf{F}_i^{[f]}(\vartheta_i^{[f]})||_F^2. \tag{4.33}$$

Estimations in (4.20) and (4.21) are similar but minimization in (4.19) is handled with a root-finding algorithm or thanks to standard numerical optimization tools considering $\partial L^{[f]}(\boldsymbol{\theta}^{[f]}, \mathbf{z}, \mathbf{x}^{[f]})$ [29], as no closed-form expression is available. More details can be found in our paper [152].

4.4 Simulation results

To highlight the robustness of the proposed MRCA and the benefit of multi-frequency consensus optimization, we compare it with the traditional Gaussian noise assumption which is addressed with the ALS method. We also compare our multi-frequency distributed scheme with the so-called mono-frequency case where (per-channel) calibration is performed for each frequency separately, without considering a specific model of variation. Each electronic gain entry $[\mathbf{g}_p^{[f]}]_k$ is generated as a complex circular Gaussian random variable with mean one and variance $\frac{1}{4}$. Ionospheric phase delay and Faraday rotation are randomly generated as function of the TEC, drawn from a uniform distribution $\mathcal{U}(1\times10^{17}, 5\times10^{17})$ and expressed in m^{-2}. We choose $N = \tilde{N} = 6$ and the Lagrangian penalty factor is fixed empirically as $\rho = 10$. Let us note that

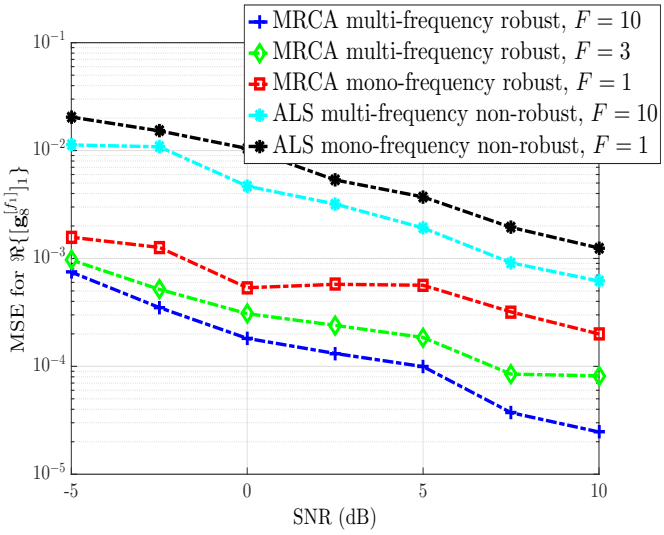


Figure 4.2: MSE of the real part of a given complex gain vs. SNR.

tuning of such parameter is addressed in [34,153]. Finally, we consider $M = 8$ receiving elements, $D = 2$ bright calibration sources and $D' = 4$ weak background sources.

We plot in Figure 4.2 the MSE of the real part of one given gain vector $\mathbf{g}^{[f_1]}$ as a function of the SNR, for different number of frequencies F which are selected in the range $75 - 125$ MHz, the behavior being the same for other parameters of interest in $\{\boldsymbol{\theta}^{[f]}\}_{f\in\mathcal{F}}$. To attain stability in convergence, less than 5 iterations are sufficient in each loop. The behavior is the same for the so-called primal and dual residuals, whose convergence mostly depends on parameters ρ, N, $\tilde{N}$ and initializations [35,36].

The estimation results show better statistical performance when calibration is performed with the proposed MRCA in comparison with state-of-the-art methods, *i.e.*, the conventional non-robust algorithm and/or the mono-frequency case.

As mentioned previously, multi-frequency robust calibration was also studied in the case of interferometric station (regime 3) with an ADMM-based distributed algorithm [152]. As in section 3.5.3, directional ionospheric phase delays are generated as linear functions of directional shifts $\boldsymbol{\alpha}_i^{[f]}$, which we aim to estimate. We still compare the MRCA^{R3} with the non-robust Gaussian case which amounts to solve a non-linear LS problem [14] and the mono-frequency case, described in Algorithm 4. The MSE for $\eta_1^{[f_1]}$ is shown as a function of the SNR in Figure 4.3 and illustrates that the multi-frequency robust calibration algorithm still provides the best statistical performance.

4.5 Conclusion

In radio astronomy, especially in regime 4, telescope design features and data processing impose major challenges in terms of calibration and the number of unknowns to estimate can be tremendously large. Thus, proposing computationally efficient algorithms is of the utmost importance and adopting distributed strategies is essential to reduce the global operational cost since the data is collected and stored among a set of computational agents. Due to separability over frequency, calibration is carried out independently for each agent and information is brought together by a fusion center which enforces consensus thanks to available constraints.

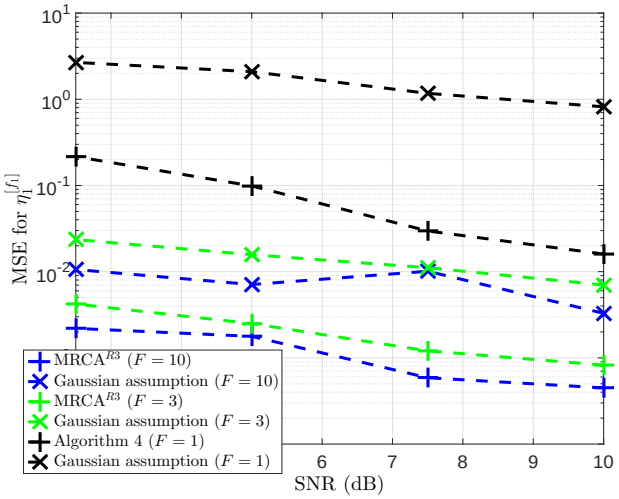


Figure 4.3: MSE of $\eta_1^{[f_1]}$ vs. SNR, for $D = 2$ bright signal sources, $M = 8$ antennas and $D' = 4$ weak background sources.

The proposed algorithm is robust and exploits frequency variation of parameters of interest in regime 4 where propagation conditions differ for each element in the array and towards each source. Such scenario is adapted for calibration of large interferometric arrays but adaptation of the multi-frequency algorithm to regime 3 is also possible. In order to go even further in the simulations and obtain reconstructed images, we will use the software MeqTrees in the next chapter to simulate realistic radio telescopes and visibility measurements [37].

Chapter 5

On realistic data simulation

5.1 Presentation of MeqTrees

In this section, we give an overview of the software MeqTrees [37], which is constituted of packages well-suited for simulation and calibration of radio astronomical data for the next generation of radio telescopes. It implements measurement equations as in (2.10) with specific Jones terms and it simulates particular radio instruments. Thus, this software system is able to build numerical models by generating realistic visibility measurements but also to solve for the unknown parameters by providing benchmark calibration algorithms. Finally, it is also possible to create images after calibration thanks to the imaging package *lwimager*. To illustrate how the software proceeds, we show a screen capture in Figure 5.1. First of all, we need to consider a Measurement Set (MS) in which all the observation setup is defined (the antenna positions, the (u, v, w) points as described in section 2.1.1, the time slots, the frequency and the pointing center). As shown in Figure 5.1, we will be considering the KAT-7 instrument, spanning 12 hours with 60 seconds of integration time per data sample, leading to 15 120 visibility points in total. We also need to define the sky model with positions and intensities of all sources. As a first step, we will be studying a set of sources taken from the Sydney University Molonglo Sky Survey (SUMSS), at 1445 MHz, using a spectral index of 0.7. A distinction is made between image-plane and uv-plane components. The first refer to the direction independent effects which occur at the antennas while the latter stand for direction dependent effects, which the third-generation calibration aims to correct as well. We notice that many types of instrumental or atmospheric disturbances can be included, including the contribution of noise.

After compilation, we define the size of data chunks being processed, which impacts the simulation speed. And we choose in which ouput column of the MS the simulated data are filled in (the MS can be viewed as a large array with many columns). This is shown in Figure 5.2. Once simulations have runned and visibilities have been generated, several parameters can be set in the imaging options: column of visibilities to consider, name of output image file, image size, ... It can be seen from (2.59) that the dirty image is a convolution between the true sky image and the dirty beam. To remove sidelobes due to imperfect uv-coverage, a deconvolution process is thus required. We notice in Figure 5.2 that the CLEAN algorithm can be specified, candidates being Clark, Hogbom and CS Clean algorithms, along with parameters describing them.

The software MeqTrees also proposes a solver in the LS sense, which minimizes the difference between observed and predicted visibilities, *i.e.*, by fitting the model to the data. To this end, the MS and the sky model, *i.e.*, the specific sources to calibrate, are required and calibration settings need to be fixed, among which the solvable parameters to consider and the type of output visibilities, as exposed in Figure 5.3. Indeed, output visibilities can be selected among

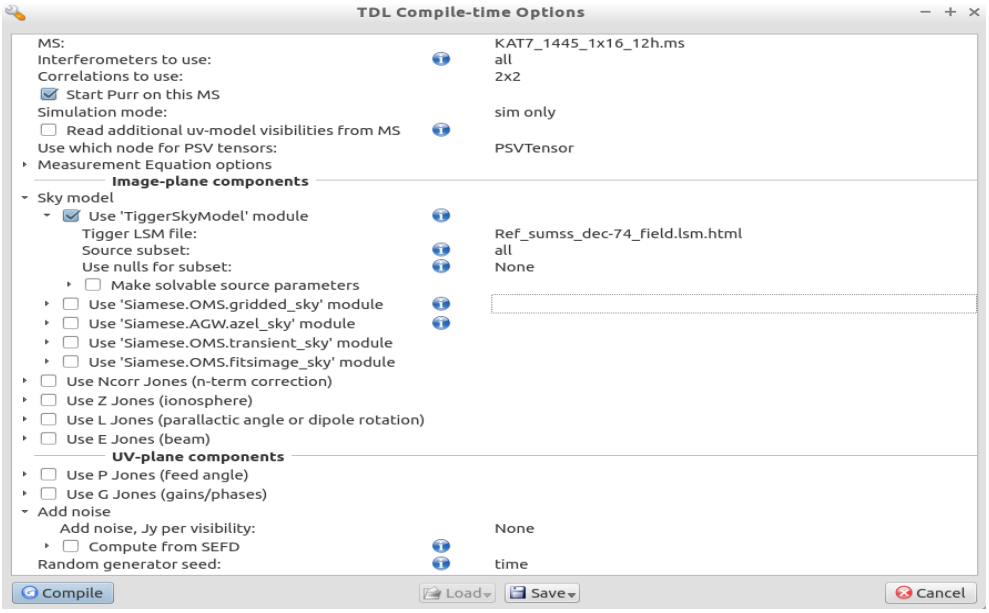


Figure 5.1: Compile-time options of the MeqTrees software for configuration of the simulated data.

Figure 5.2: Runtime options of the MeqTrees software for running the simulation and performing imaging.

the corrected data, the (un)corrected residuals and the predict. Let us consider calibration along the i-th source and for the antenna pair (p, q). If we note $\mathbf{X}_{pq}$ the available data, *i.e.*, the observed visibilities, and consider the following corrupted model (also called the predict)

$$\mathbf{V}_{i,pq}(\boldsymbol{\theta}) = \mathbf{G}_p(\mathbf{g}_p)\mathbf{K}_{i,p}\mathbf{C}_i\mathbf{K}_{i,q}^H\mathbf{G}_q^H(\mathbf{g}_q) \tag{5.1}$$

with only the typical propagation delay $\mathbf{K}_{i,p}$ Jones matrix and the direction inpendent electronic gain $\mathbf{G}_p$ term, then calibration amounts to minimize $\mathbf{X}_{pq} - \mathbf{V}_{i,pq}(\boldsymbol{\theta})$ *w.r.t.* $\boldsymbol{\theta}$ which includes the

per-antenna electronic gains $\mathbf{g}_p$ and $\mathbf{g}_q$. The corrected and uncorrected residuals are, respectively, given by

$$\hat{\mathbf{G}}_p^{-1}(\mathbf{X}_{pq} - \mathbf{V}_{i,pq}(\hat{\boldsymbol{\theta}}))\hat{\mathbf{G}}_q^{H^{-1}} \tag{5.2}$$

and

$$\mathbf{X}_{i,pq} - \mathbf{V}_{i,pq}(\hat{\boldsymbol{\theta}}). \tag{5.3}$$

Finally, the corrected data are written as

$$\hat{\mathbf{G}}_p^{-1}\mathbf{X}_{pq}\hat{\mathbf{G}}_q^{H^{-1}}. \tag{5.4}$$

Considering the residuals enables to remove the contribution of known reference sources and unveil the weak unknown signals buried in noise and bright foregrounds. Thus, accurate calibration leads to better substraction of bright sources from the observed data and enables imaging of faint background sources which are hardly visible otherwise. We will pay particular attention to this kind of output visibilities.

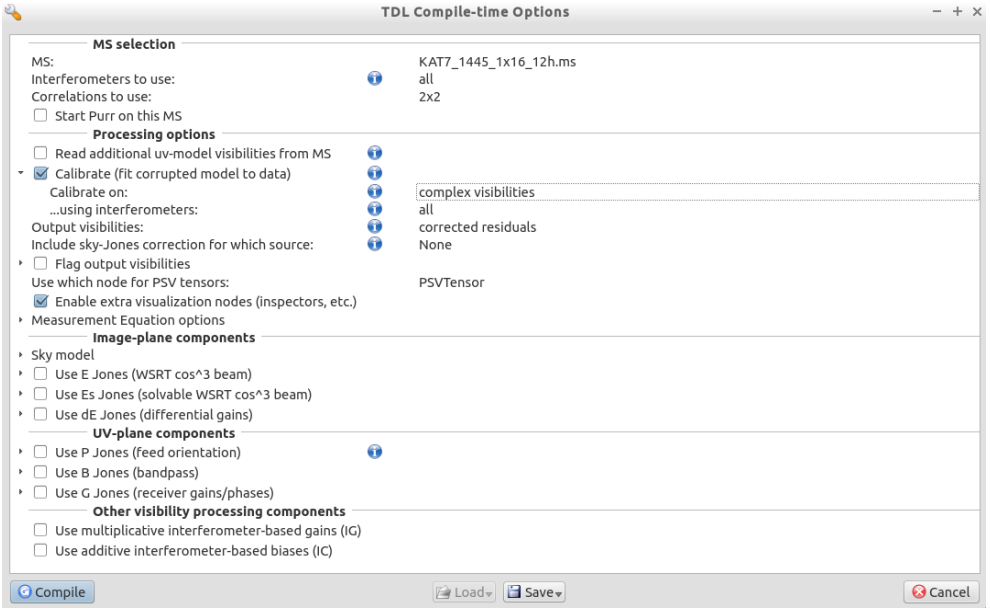


Figure 5.3: Compile-time options of the MeqTrees software for calibration solver.

5.2 Handling of MeqTrees: academic examples

To begin with, we start with a selected set of sources from the SUMSS, assumed to be unpolarized. Thus, only the intensity I is non-zero in the source brightness introduced in (2.10). They are exposed in Table 5.1 and we will refer to them thanks to the two first numbers in the name.

Table 5.1: Selection of point-like sources from the SUMMS.

Name	RA (hours h, minutes m, seconds s)	DEC (degrees °, minutes ', seconds ")	Intensity (Jansky Jy)
38_$J1845M76$	19 12 16.78	− 75 34 28.10	0.1097
43_$J1930M76$	19 15 54.44	− 74 39 36.90	1.142

In Figure 5.4, we expose the benefit of deconvolution by showing the dirty (left) and the clean (right) images, with the single bright source number 43 in the middle which is free of any

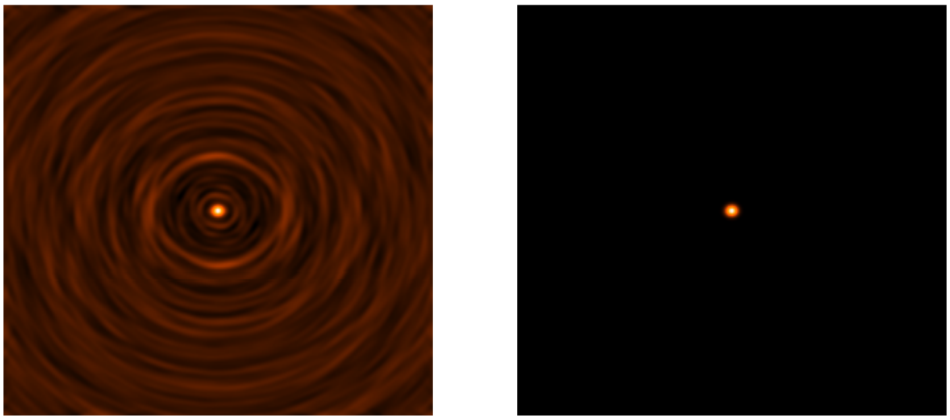


Figure 5.4: Dirty (left) and clean (right) images in a corruption-free environment and a single source in the sky model.

disturbances here. Likewise, let us consider a diffuse radio source, named M31, where Crab-like supernova remnants are located. To recover such complex structure, fine measurements are required, thus, in this specific case, we consider the MeerKAT radio telescope whose dense uv coverage was shown in Figure 2.2. Indeed, by considering 64 antennas during 4 hours and 60 seconds integration time per data sample, the full number of visibility points is 483 840 instead of 15 120. In Figure 5.5, we expose the true ideal image along with the clean and dirty images. In the rest of the chapter, we stick to the KAT-7 instrument for computational issues.

To highlight the effects due to the presence of unknown electronic receiver gains, we make the dirty images with and without simulating periodically varying errors for gains $\mathbf{G}$. Visual results are exposed in Figure 5.6, where the bright source number 43 and the weaker source number 38 are considered. The position of the latter is denoted thanks to a red square around it, in the uncorrupted case. We notice that source number 38 is hardly visible but becomes fully hidden when electronic gains are included in the generated data. However, as mentioned before, it is possible to reveal the faint background by performing accurate estimation of perturbations and computing residuals. For that purpose, we still consider the previously introduced $\mathbf{G}$ Jones terms and perform calibration along source number 43 with the solver in MeqTrees. Dirty images associated to uncorrected and corrected residuals are shown in Figure 5.7. As expected, recovering of weak source number 38 is possible.

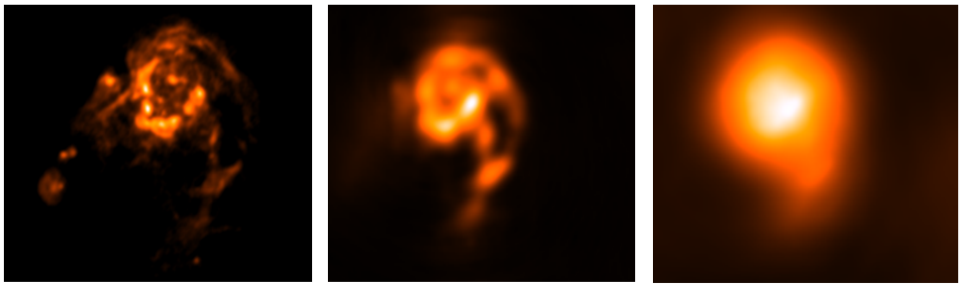


Figure 5.5: Ideal (left), clean (middle) and dirty (right) images of the diffuse M31 source without any perturbations.

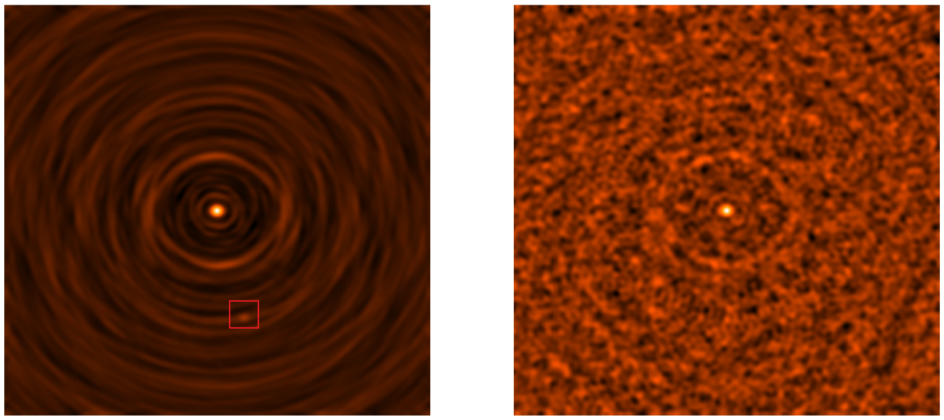


Figure 5.6: Dirty images with (right) and without (left) considering disruptive **G** Jones gains.

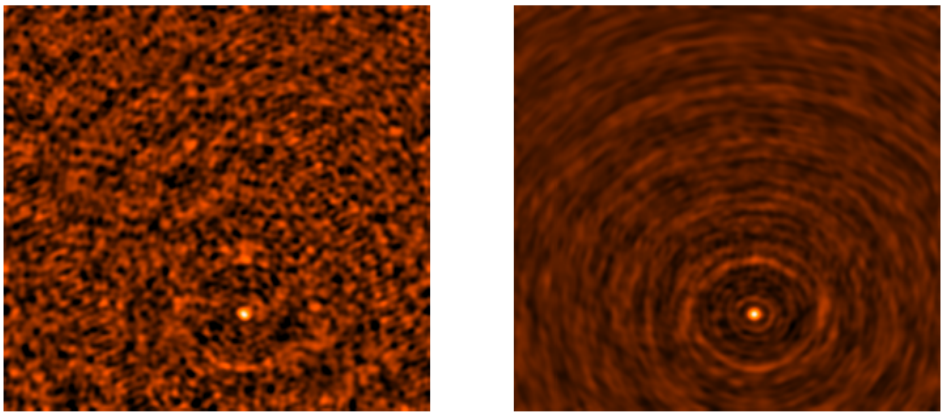


Figure 5.7: Dirty images associated to the uncorrected residuals (left) and corrected residuals (right), solving for **G** Jones terms.

In the following, we aim to compare the proposed algorithms, presented in the previous chapters, with the MeqTrees solver.

5.3 Comparison with the proposed algorithms

We wish to test the designed robust calibration techniques on realistic data sets simulated with MeqTrees and evaluate the results directly on the reconstructed residual images. To achieve this, the following steps are required

- Simulation of realistic corrupted data under MeqTrees
- Retrieve the generated data and perform calibration with the proposed algorihms
- Insert the output calibrated visibilities, *e.g.*, the corrected residuals, into the MS
- Make the corresponding dirty or clean images with *lwimager*

Table 5.2: Study of $D' = 16$ weak sources from the SUMMS.

Name	RA (h,m,s)	DEC (°, ', ")	Intensity (Jy)
16_$J1845M76$	19 01 34.74	− 74 44 01.40	0.0602
17_$J1930M76$	19 01 39.30	− 74 47 24.00	0.0998
18_$J1930M76$	19 02 18.01	− 74 04 24.10	0.0632
23_$J1930M72$	19 05 48.63	− 73 49 48.20	0.1206
25_$J1845M76$	19 06 32.35	− 74 22 08.70	0.0816
26_$J1845M76$	19 06 56.66	− 75 11 00.30	0.0826
31_$J1845M76$	19 10 22.68	− 74 17 36.60	0.1167
33_$J1845M76$	19 10 46.30	− 74 55 32.80	0.1059
34_$J1930M76$	19 10 50.15	− 74 23 49.80	0.0699
38_$J1845M76$	19 12 16.78	− 75 34 28.10	0.1097
41_$J1845M76$	19 14 10.38	− 75 15 41.90	0.066
47_$J1930M76$	19 17 33.14	− 74 32 59.20	0.0746
49_$J1930M76$	19 18 33.39	− 74 45 17.40	0.0774
50_$J1930M72$	19 19 02.38	− 74 02 50.30	0.0737
54_$J1930M76$	19 21 14.53	− 74 25 59.60	0.0583
59_$J1930M76$	19 26 48.44	− 73 58 19.20	0.0575

Let us consider calibration source number 43 and $D' = 16$ weak sources taken from the SUMSS, described in Table 5.2. They are also illustrated in Figure 5.8, without any noise or perturbation effects.

We assume that all sources, calibration and non-calibration ones, are corrupted by Faraday rotation matrices, noise is also added and we compare the calibration solver in MeqTrees with the previously introduced MRCA^{R3}. In Figure 5.9, we show the corrected residual images at 895 MHz in a small area surrounding the calibration source number 43, whose position is indicated by a red cross. We notice better removal of the calibration source for the proposed multi-frequency robust calibration algorithm while weak sources number 47 and 49 are also visible on the left hand side of the zoomed image. To illustrate how well a weak background source is recovered, we plot in Figure 5.10 the peak flux of weak source number 41 (non-calibration background source) as a function of frequency. Better flux recovering is attained with the MRCA^{R3}.

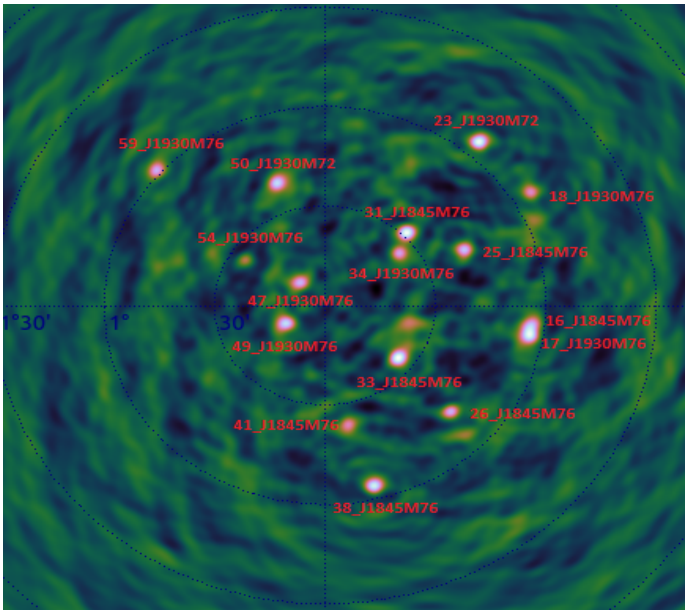


Figure 5.8: Uncorrupted dirty image with the $D' = 16$ weak sources we wish to recover after calibration.

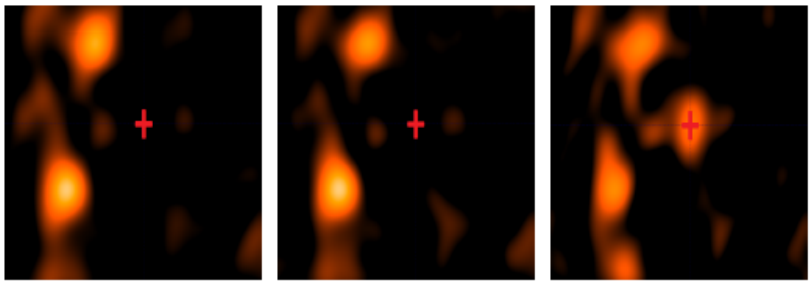


Figure 5.9: Corrected residual images for the ideal case (left), the proposed MRCA^{R3} (middle) and the calibration solver in MeqTrees (right).

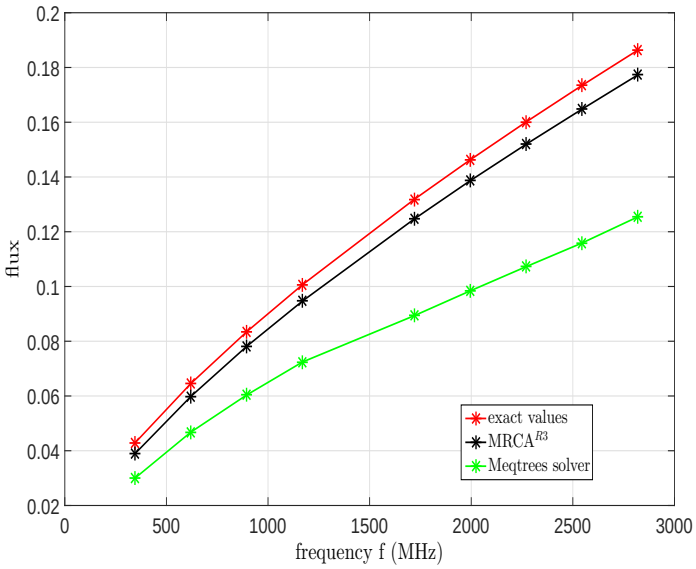
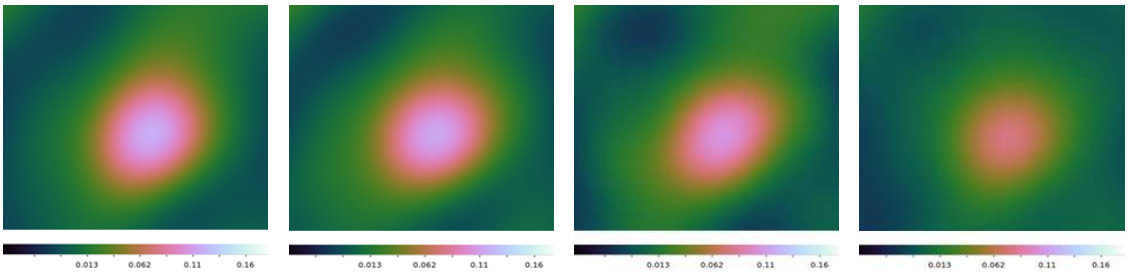


Figure 5.10: Recovered flux of one weak background source after calibration for $F = 9$ frequencies.

Furthermore, we wish to show the performance of the MRCA, described in Algorithm 5, which is adapted for calibration of large interferometric arrays in regime 4. In Figures 5.11 and 5.12, we still represent the corrected residual images in restricted areas surrounding specific background sources, once calibration and subtraction of the bright source have been performed. In Figure 5.11, the peak flux intensity of the recovered source (number 59) is respectively given by (a) true 0.11856 Jy, (b) 0.115736 Jy, (c) 0.105937 Jy and (d) 0.08049 Jy. The corresponding recovered peak flux of the three sources in Figure 5.12 is given in Table 5.3.

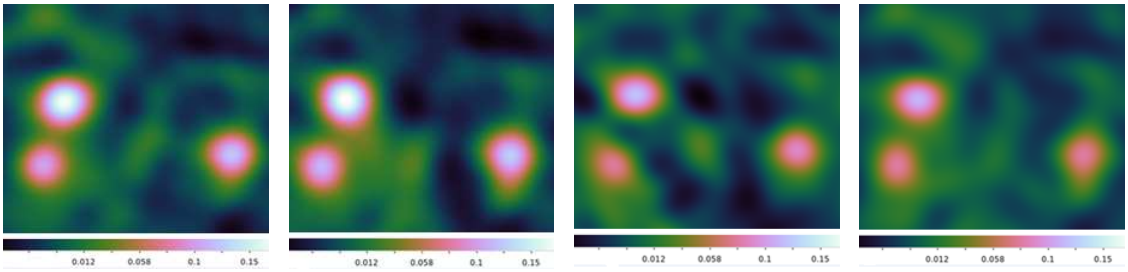
Table 5.3: Recovered peak flux (in Jy) for sources in Figure 5.12.

Name	Source number 31	Source number 34	Source number 25
(a) True	0.167113	0.108398	0.122119
(b)	0.165512	0.107406	0.122114
(c)	0.121019	0.078651	0.0897499
(d)	0.116302	0.0781593	0.0815132



(a) (b) (c) (d)

Figure 5.11: Corrected residual for (a) ideal case, (b) MRCA, (c) multi-frequency non-robust calibration and (d) mono-frequency non-robust calibration, around the weakest background source, *i.e.*, source number 59.



(a) (b) (c) (d)

Figure 5.12: Corrected residual for (a) ideal case, (b) MRCA (c) multi-frequency non-robust calibration and (d) mono-frequency non-robust calibration, around three background source, *i.e.*, sources number 25, 31 and 34.

Let us note that ionospheric perturbation effects corrupt the data here. By multi-frequency non-robust calibration, we mean that a distributed strategy is adopted to handle multiple frequencies, *e.g.*, with the ADMM procedure, but the algorithm is based on a Gaussian noise assumption. As regards the calibration solver in MeqTrees, we also name it mono-frequency non-robust calibration as we do not exploit information across frequency (calibration is performed for each frequency separately) and the Gaussian noise model is typically assumed.

5.4 Conclusion

The MeqTrees software system is uniquely suited for generation of realistic radio astronomical data and also proposes benchmark calibration algorithms, adapted to new radio telescopes. Various direction dependent and independent effects can be incorporated such as the receiver complex gains, the frequency bandpass and the primary beam of antennas. Each individual effect is viewed as a separable term in the software and can be corrected independently during calibration. Dirty and clean images can also be created in order to reveal the benefit of calibration and the presence of weak sources of interest.

During my stay at Lab. J.-L. Lagrange in Nice (with Prof. A. Ferrari), work was conducted to master the essential features of the MeqTrees system (data simulation, calibration and imaging) and secure the linkage between the designed robust calbiration algorithms from the *Matlab* environment and this software widely used by the community of researchers in radio astronomy. This was not straightforward and has taken time to develop *Python* scripts which handle the

MS data sets, to study the robustness aspects by uncluding the CG-based approaches and extend to the multi-frequency scenario. The flexibility of the software also makes it possible to include our own simulated perturbation effects on the basic uncorrupted data, generated by the interferometer under study. In brief, the software MeqTrees enables to understand better what an interferometer would measure in practice but also to consider the practical feasibility of the proposed techniques in this document.

Chapter 6

Conclusion and prospects

The work described in this document deals with calibration of radio interferometers in non-Gaussian environment and multi-frequency scenario. Specifically, a statistical study revealed that the classical Gaussian noise assumption was not suitable to model the contribution of outliers in visibility measurements, which are due to weak unknown background sources or radio frequency interferences. To address this problem, we considered the CG distribution, composed of texture and speckle parts. Deterministic and Bayesian approaches have been proposed but numerical simulations emphasized that the relaxed version (deterministic approach) was more flexible and avoided model misspecifications. Specifically, this latter approach is based on the ML scheme. It enables to estimate iteratively the noise parameters, and more importantly, the Jones matrices which account for all perturbation effects along the signal propagation path. Such matrices can be viewed as non-structured or structured during the estimation process. To reduce the computational burden, the EM and the BCD algorithms were used, leading to closed-form expressions. For comparative purposes, traditional calibration algorithms as the classical Gaussian noise assumption and the robust Student's t-based method were considered.

To deal with multiple frequencies, we made use of a distributed and consensus strategy by exploiting the specific variation of parameters of interest *w.r.t.* frequency. Some variation models are available in the literature as it is the case for ionospheric disruptions, otherwise smoothness across frequency can be assumed using polynomials. To perform consensus optimization, the ADMM procedure was employed and the global computational load was distributed across a network of agents while the fusion center enforced consensus. All along the study, we focused on two particular regimes, *i.e.*, regimes 3 and 4 as direction dependent distortions are taken into account and such scenarios are more representative of the new generation of radio interferometers. The designed multi-frequency robust calibration algorithm was assessed against state-of-the art mono-frequency and/or non robust cases.

Finally, the use of the sotftware MeqTrees enabled to simulate realistic radio data sets with specified perturbation effects and to provide a benchmark calibration algorithm that has been compared to our work. During the calibration process, weak unknown sources contribute to the noise, affecting and changing its classical Gaussian nature but during imaging, they become the elements of interest we wish to reveal as they are not readily accessible. Thus, a better removal of calibration sources and recovering of background sources in residual images indicates a more accurate estimation of perturbation effects during calibration and enables to evaluate the performances of the designed algorithms.

Among the possible future prospects which naturally follow this work, we can mention the following points

- The temporal variation of parameters of interest could be included into the calibration process [73]. Proposing independent solutions between blocks of time intervals is a straightforward extension and only requires to add a sum over time in the likelihood expression. However, solving for fast- or slow-varying effects *w.r.t.* time, thus considering different time scales remains to be done. A similar approach to the one proposed in the multi-frequency scenario can be adopted.

- Compressive sensing approaches are used increasingly in radio astronomy imaging as they promote sparsity (mostly empty sky) and lead to good reconstruction results, providing competitive results with respect to traditional methods as CLEAN [99,123,154,155]. However, the Gaussian noise assumption is usually still considered. Thus, combining both aspects (sparsity and robustness) remains to be developed in future work.

- Optimization on matrix manifolds has developed significantly recently and yields better performing algorithms when the geometric structure is exploited [156]. Calibration is essentially an optimization problem and reformulation of the problem on a Riemannian manifold has already been proposed for radio interferometric calibration and leads to faster convergence with reduced computational cost compared to traditional calibration on Euclidean space [157]. Therefore, the designed robust algorithms in this work could benefit from adaptation to Riemannian manifold, using the RTR method for instance [115].

- The CRB is of crucial importance in estimation theory and statistics as it provides the minimal variance of unbiased estimators [145,158]. But it can also be considered for design configuration issues in antennas of next stations [159–161]. Indeed, the patterning can affect the performances of deconvolution and calibration processes and studying performance bounds can give an idea about the optimal design of stations as regards the number of antennas to consider, their positions (randomized or following a specific geometry) and the array aperture. Furthermore, studying the CRB is of particular interest, notably in order to investigate the impact of misspecified models (noise distribution, frequency variation of parameters, antenna positions) [162].

- Testing the designed algorithms on real data is worth exploring and requires adaptation to very large scales [6, 53]. Indeed, with a tremendous number of radio antennas and visible radio sources in their field-of-view, the computational cost can become a restricting factor. Efforts have been made in this work to propose algorithms with low computational complexity but deeper cost-efficiency analysis would facilitate this future test phase.

Appendix A

Per-baseline speckle component in the CG distribution

We describe here the corresponding expressions of (3.27) and (3.29) when a different speckle matrix $\mathbf{\Omega}_{pq}$ is assumed for each baseline $p < q,\ \ p, q \in \{1, \ldots, M\}^2$. In this case, the log-likelihood function is written as

$$\begin{aligned} \ln p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\tau}, \mathbf{\Omega}_{12}, \mathbf{\Omega}_{13}, \ldots, \mathbf{\Omega}_{(M-1)M}) = &- 4B \ln \pi - 4 \sum_{pq} \ln \tau_{pq} - \sum_{pq} \ln |\mathbf{\Omega}_{pq}| \\ &- \sum_{pq} \frac{1}{\tau_{pq}} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \mathbf{\Omega}_{pq}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}). \end{aligned} \tag{A.1}$$

For each antenna pair, the texture estimate reads

$$\hat{\tau}_{pq} = \frac{1}{4} \mathbf{u}_{pq}^H(\boldsymbol{\theta}) \mathbf{\Omega}_{pq}^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta}) \tag{A.2}$$

while the per-baseline speckle component is given by

$$\hat{\mathbf{\Omega}}_{pq}^{t+1} = 4 \frac{\mathbf{u}_{pq}(\boldsymbol{\theta}) \mathbf{u}_{pq}^H(\boldsymbol{\theta})}{\mathbf{u}_{pq}^H(\boldsymbol{\theta}) \left(\hat{\mathbf{\Omega}}_{pq}^t \right)^{-1} \mathbf{u}_{pq}(\boldsymbol{\theta})}. \tag{A.3}$$

The remainder of the algorithm is straightforwardly obtained using (A.3).

Appendix B

Use of the BCD algorithm for estimation of non-structured Jones matrices

We present here the main steps leading to (3.43). Firstly, for sake of clarity, let us denote $\mathbf{c}_i = [c_{i_1}, c_{i_2}, c_{i_3}, c_{i_4}]^T$ to refer to the four entries of the vectorization of source coherency matrix $\mathbf{C}_i$. Likewise, for the i-th source, p-th antenna and q-th antenna, respectively, we write

$$\mathbf{J}_{i,p}(\boldsymbol{\theta}_{i,p}) = \begin{bmatrix} p_{i_1} & p_{i_2} \\ p_{i_3} & p_{i_4} \end{bmatrix} \quad \text{and} \quad \mathbf{J}_{i,q}(\boldsymbol{\theta}_{i,q}) = \begin{bmatrix} q_{i_1} & q_{i_2} \\ q_{i_3} & q_{i_4} \end{bmatrix} \tag{B.1}$$

, i.e., $\boldsymbol{\theta}_{i,p} = [p_{i_1}, p_{i_2}, p_{i_3}, p_{i_4}]^T$ and $\boldsymbol{\theta}_{i,q} = [q_{i_1}, q_{i_2}, q_{i_3}, q_{i_4}]^T$.
Using these latter notation, we obtain (3.41) where

$$\boldsymbol{\Sigma}_{i,q} = \begin{bmatrix} \alpha_{i,q} & \beta_{i,q} & 0 & 0 \\ 0 & 0 & \alpha_{i,q} & \beta_{i,q} \\ \gamma_{i,q} & \rho_{i,q} & 0 & 0 \\ 0 & 0 & \gamma_{i,q} & \rho_{i,q} \end{bmatrix} \tag{B.2}$$

in which $\alpha_{i,q} = q_{i_1}^* c_{i_1} + q_{i_2}^* c_{i_3}$, $\beta_{i,q} = q_{i_1}^* c_{i_2} + q_{i_2}^* c_{i_4}$, $\gamma_{i,q} = q_{i_3}^* c_{i_1} + q_{i_4}^* c_{i_3}$ and $\rho_{i,q} = q_{i_3}^* c_{i_2} + q_{i_4}^* c_{i_4}$.
We also obtain (3.42) where

$$\boldsymbol{\Upsilon}_{i,q} = \begin{bmatrix} \lambda_{i,q} & \mu_{i,q} & 0 & 0 \\ \nu_{i,q} & \xi_{i,q} & 0 & 0 \\ 0 & 0 & \lambda_{i,q} & \mu_{i,q} \\ 0 & 0 & \nu_{i,q} & \xi_{i,q} \end{bmatrix} \tag{B.3}$$

in which $\lambda_{i,q} = q_{i_1} c_{i_1} + q_{i_2} c_{i_2}$, $\mu_{i,q} = q_{i_1} c_{i_3} + q_{i_2} c_{i_4}$, $\nu_{i,q} = q_{i_3} c_{i_1} + q_{i_4} c_{i_2}$ and $\xi_{i,q} = q_{i_3} c_{i_3} + q_{i_4} c_{i_4}$.

Finally, the cost function in (3.40) can be written as

$$\begin{aligned} \phi_i(\boldsymbol{\theta}_{i,p}) = &\Big(\mathbf{w}_{i,p} - \mathbf{s}_{i,p}(\boldsymbol{\theta}_{i,p})\Big)^H \mathbf{A}_{i,p} \Big(\mathbf{w}_{i,p} - \mathbf{s}_{i,p}(\boldsymbol{\theta}_{i,p})\Big) + \\ &\Big(\tilde{\mathbf{w}}_{i,p} - \tilde{\mathbf{s}}_{i,p}(\boldsymbol{\theta}_{i,p})\Big)^H \tilde{\mathbf{A}}_{i,p} \Big(\tilde{\mathbf{w}}_{i,p} - \tilde{\mathbf{s}}_{i,p}(\boldsymbol{\theta}_{i,p})\Big) + \text{Constant} \end{aligned} \tag{B.4}$$

where

- $\mathbf{w}_{i,p} = [\mathbf{w}_{i,p(p+1)}^T, \ldots, \mathbf{w}_{i,pM}^T]^T$

- $\mathbf{s}_{i,p}(\boldsymbol{\theta}_{i,p}) = [\mathbf{s}^T_{i,p(p+1)}(\boldsymbol{\theta}_{i,p}), \ldots, \mathbf{s}^T_{i,pM}(\boldsymbol{\theta}_{i,p})]^T$
- $\mathbf{A}_{i,p} = \text{bdiag}\{\beta_i\tau_{p(p+1)}\boldsymbol{\Omega}, \ldots, \beta_i\tau_{pM}\boldsymbol{\Omega}\}^{-1}$.

Furthermore, we have

- $\tilde{\mathbf{w}}_{i,p} = [\mathbf{w}^{*T}_{i,1p}, \ldots, \mathbf{w}^{*T}_{i,(p-1)p}]^T$
- $\tilde{\mathbf{s}}_{i,p}(\boldsymbol{\theta}_{i,p}) = [\mathbf{s}^{*T}_{i,1p}(\boldsymbol{\theta}_{i,p}), \ldots, \mathbf{s}^{*T}_{i,(p-1)p}(\boldsymbol{\theta}_{i,p})]^T$
- $\tilde{\mathbf{A}}_{i,p} = \text{bdiag}\{\beta_i\tau_{1p}\boldsymbol{\Omega}^*, \ldots, \beta_i\tau_{(p-1)p}\boldsymbol{\Omega}^*\}^{-1}$.

We make use of (3.41) in what follows

$$\mathbf{s}_{i,p}(\boldsymbol{\theta}_{i,p}) = \begin{bmatrix} \mathbf{s}_{i,p(p+1)}(\boldsymbol{\theta}_{i,p}) \\ \vdots \\ \mathbf{s}_{i,pM}(\boldsymbol{\theta}_{i,p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{i,p+1}\boldsymbol{\theta}_{i,p} \\ \vdots \\ \boldsymbol{\Sigma}_{i,M}\boldsymbol{\theta}_{i,p} \end{bmatrix} = \boldsymbol{\Sigma}_i\boldsymbol{\theta}_{i,p} \tag{B.5}$$

where $\boldsymbol{\Sigma}_i = [\boldsymbol{\Sigma}^T_{i,p+1}, \cdots, \boldsymbol{\Sigma}^T_{i,M}]^T$. Likewise, we use (3.42) in

$$\tilde{\mathbf{s}}_{i,p}(\boldsymbol{\theta}_{i,p}) = \begin{bmatrix} \mathbf{s}^*_{i,1p}(\boldsymbol{\theta}_{i,p}) \\ \vdots \\ \mathbf{s}^*_{i,(p-1)p}(\boldsymbol{\theta}_{i,p}) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Upsilon}^*_{i,1}\boldsymbol{\theta}_{i,p} \\ \vdots \\ \boldsymbol{\Upsilon}^*_{i,p-1}\boldsymbol{\theta}_{i,p} \end{bmatrix} = \boldsymbol{\Upsilon}_i\boldsymbol{\theta}_{i,p} \tag{B.6}$$

in which $\boldsymbol{\Upsilon}_i = [\boldsymbol{\Upsilon}^{*T}_{i,1}, \cdots, \boldsymbol{\Upsilon}^{*T}_{i,p-1}]^T$.

Inserting (B.5) and (B.6) into (B.4) and taking the derivative w.r.t. $\boldsymbol{\theta}_{i,p}$ leads to the expressions in (3.43), using the fact that $\mathbf{A}_{i,p}$ and $\tilde{\mathbf{A}}_{i,p}$ are Hermitian.

Appendix C

Estimation of the ionospheric phase delays in multi-frequency scenario

We specify here the notations which lead to (4.24). To this end, we need to consider $\frac{\partial l^{[f]}(\boldsymbol{\theta}^{[f]})}{\partial \varphi_{i,p}^{[f]}}$ and $\frac{\partial h^{[f]}(\boldsymbol{\theta}^{[f]})}{\partial \varphi_{i,p}^{[f]}}$. Thus, minimization (4.19) *w.r.t.* phase delay $\varphi_{i,p}^{[f]}$ leads to

$$\beta_{i,p}^{[f]} \exp\left(j\varphi_{i,p}^{[f]}\right) + \alpha_{i,p}^{[f]} \exp\Big(-j\varphi_{i,p}^{[f]}\Big) = 0 \tag{C.1}$$

in which

- $\beta_{i,p}^{[f]} = \sum_{\substack{q=1 \\ q>p}}^{M} \frac{j}{\tau_{pq}^{[f]}} \exp\Big(-j\varphi_{i,q}^{[f]}\Big)\delta_{i,pq}^{[f]} + \sum_{\substack{q=1 \\ q<p}}^{M} \frac{j}{\tau_{qp}^{[f]}} \exp\Big(-j\varphi_{i,q}^{[f]}\Big)\delta_{i,qp}^{[f]^H} + jx_{i,p}^{[f]^*} - j\rho \mathbf{z}_{i,p}^H \mathbf{b}^{[f]^*}$
- $\alpha_{i,p}^{[f]} = \sum_{\substack{q=1 \\ q>p}}^{M} \frac{-j}{\tau_{pq}^{[f]}} \exp\left(j\varphi_{i,q}^{[f]}\right)\delta_{i,pq}^{[f]^H} - \sum_{\substack{q=1 \\ q<p}}^{M} \frac{j}{\tau_{qp}^{[f]}} \exp\left(j\varphi_{i,q}^{[f]}\right)\delta_{i,qp}^{[f]} - jx_{i,p}^{[f]} + j\rho \mathbf{b}^{[f]^T} \mathbf{z}_{i,p}.$

We also introduce

- $\delta_{i,pq}^{[f]} = \left(-\mathbf{v}_{pq}^{[f]^H} + \sum_{\substack{k=1 \\ k\neq i}}^{D} \mathbf{s}_{k,pq}^{[f]^H}\right) \boldsymbol{\Omega}^{[f]^{-1}} \mathbf{d}_{i,pq}^{[f]}$
- $\mathbf{d}_{i,pq}^{[f]} = \left(\mathbf{G}_q^{[f]^*}(\mathbf{g}_q^{[f]})\mathbf{H}_{i,q}^{[f]^*}\mathbf{F}_{i,q}^{[f]}(\vartheta_{i,q}^{[f]})\right) \otimes \left(\mathbf{G}_p^{[f]}(\mathbf{g}_p^{[f]}) \otimes \mathbf{H}_{i,p}^{[f]}\mathbf{F}_{i,p}^{[f]}(\vartheta_{i,p}^{[f]})\right) \mathbf{c}_i^{[f]}.$

Finally, we directly deduce (4.24) from (C.1).

Appendix D

Estimation of the complex electronic gains in multi-frequency scenario

We describe here the notations introduced in the estimation of frequency dependent electronic gains. The expressions in (4.26) are given by

$$\lambda_{pq}^{[f]} = t_{pq}^{[f]} + \mathbf{w}_{pq}^{[f]^H} \left(\mathbf{I}_2 \otimes \mathbf{G}_p^{[f]^*} \right) \boldsymbol{\varpi}_{pq}^{[f]} \tag{D.1}$$

in which

- $t_{pq}^{[f]} = -\mathbf{v}_{pq}^{[f]^H} \mathbf{m}_{pq}^{[f]}$
- $\mathbf{m}_{pq}^{[f]} = \boldsymbol{\Omega}^{[f]^{-1}} \left(\mathbf{G}_q^{[f]^*} \otimes \mathbf{E}_1 \right) \mathbf{w}_{pq}^{[f]}$
- $\mathbf{E}_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$.

Besides, $\mathbf{w}_{pq}^{[f]} = \sum_{i=1}^{D} \left(\mathbf{H}_{i,q}^{[f]^*} \mathbf{Z}_{i,q}^{[f]^*}(\varphi_{i,q}) \mathbf{F}_{i,q}^{[f]}(\vartheta_{i,q}^{[f]}) \right) \otimes \left(\mathbf{H}_{i,p}^{[f]} \mathbf{Z}_{i,p}^{[f]}(\varphi_{i,p}) \mathbf{F}_{i,p}^{[f]}(\vartheta_{i,p}^{[f]}) \right) \mathbf{c}_i^{[f]}$ and $\boldsymbol{\varpi}_{pq}^{[f]} = \left(\mathbf{G}_q^{[f]} \otimes \mathbf{I}_2 \right) \mathbf{m}_{pq}^{[f]}$.

Similarly, we have

$$\tilde{\lambda}_{qp}^{[f]} = \tilde{t}_{qp}^{[f]} + \tilde{\mathbf{w}}_{qp}^{[f]^H} \left(\mathbf{G}_p^{[f]^*} \otimes \mathbf{I}_2 \right) \tilde{\boldsymbol{\varpi}}_{qp}^{[f]} \tag{D.2}$$

in which

- $\tilde{t}_{qp}^{[f]} = -\tilde{\mathbf{w}}_{qp}^{[f]^H} \mathbf{v}_{qp}^{[f]}$
- $\tilde{\mathbf{w}}_{qp}^{[f]} = \boldsymbol{\Omega}^{[f]^{-1}} \left(\mathbf{E}_1 \otimes \mathbf{G}_q^{[f]} \right) \mathbf{w}_{qp}^{[f]}$,
- $\tilde{\boldsymbol{\varpi}}_{qp}^{[f]} = \left(\mathbf{I}_2 \otimes \mathbf{G}_q^{[f]} \right) \mathbf{w}_{qp}^{[f]}$.

The derivative in (4.26) can be written more compactly

$$\frac{\partial l^{[f]}(\mathbf{g}^{[f]})}{\partial [\mathbf{g}_p^{[f]}]_1} = t_p^{[f]} + \mathbf{w}_p^{[f]^H} \left(\mathbf{I}_S \otimes \mathbf{I}_2 \otimes \mathbf{G}_p^{[f]^*} \right) \boldsymbol{\omega}_p^{[f]} + \tilde{\mathbf{w}}_p^{[f]^H} \left(\mathbf{I}_V \otimes \mathbf{G}_p^{[f]^*} \otimes \mathbf{I}_2 \right) \tilde{\boldsymbol{\omega}}_p^{[f]} \tag{D.3}$$

in which

- $t_p^{[f]} = \sum_{\substack{q=1 \\ q>p}}^{M} \frac{1}{\tau_{pq}^{[f]}} t_{pq}^{[f]} + \sum_{\substack{q=1 \\ q<p}}^{M} \frac{1}{\tau_{qp}^{[f]}} \tilde{t}_{qp}^{[f]}$,
- $\mathbf{w}_p^{[f]} = [\frac{1}{\tau_{p(p+1)}^{[f]}} \mathbf{w}_{p(p+1)}^{[f]^T}, \ldots, \frac{1}{\tau_{pM}^{[f]}} \mathbf{w}_{pM}^{[f]^T}]^T$
- $\tilde{\mathbf{w}}_p^{[f]} = [\frac{1}{\tau_{1p}^{[f]}} \tilde{\mathbf{w}}_{1p}^{[f]^T}, \ldots, \frac{1}{\tau_{(p-1)p}^{[f]}} \tilde{\mathbf{w}}_{(p-1)p}^{[f]^T}]^T$.

Similarly, we introduce $\boldsymbol{\omega}_p^{[f]} = [\boldsymbol{\varpi}_{p(p+1)}^{[f]^T}, \ldots, \boldsymbol{\varpi}_{pM}^{[f]^T}]^T$ and $\tilde{\boldsymbol{\omega}}_p^{[f]} = [\tilde{\boldsymbol{\varpi}}_{1p}^{[f]^T}, \ldots, \tilde{\boldsymbol{\varpi}}_{(p-1)p}^{[f]^T}]^T$.

If we are considering the p-th array element, we note $S = (M - p)$ and $V = (p - 1)$. Let us define the permutation matrix $\mathbf{P} = [\mathbf{e}_1, \mathbf{e}_3, \mathbf{e}_2, \mathbf{e}_4]$ where the 4×1 vector $\mathbf{e}_k$ has zeros except at the k-th position which is equal to unity. With $\boldsymbol{\varsigma}_p^{[f]} = \left(\mathbf{I}_V \otimes \mathbf{P}^T\right) \tilde{\mathbf{w}}_p^{[f]}$ and $\boldsymbol{\varrho}_p^{[f]} = (\mathbf{I}_V \otimes \mathbf{P}) \, \tilde{\boldsymbol{\omega}}_p^{[f]}$, we obtain the following one-dimensional linear equation

$$\begin{aligned} \frac{\partial l^{[f]}(\mathbf{g}^{[f]})}{\partial [\mathbf{g}_p^{[f]}]_1} =& t_p^{[f]} + [\mathbf{g}_p^{[f]^*}]_1 \sum_{k=0}^{2S-1} [\mathbf{w}_p^{[f]^*}]_{2k+1} [\boldsymbol{\omega}_p^{[f]}]_{2k+1} + [\mathbf{g}_p^{[f]^*}]_1 \sum_{k=0}^{2V-1} [\boldsymbol{\varsigma}_p^{[f]^*}]_{2k+1} [\boldsymbol{\varrho}_p^{[f]}]_{2k+1} \\ &+ [\mathbf{g}_p^{[f]^*}]_2 \sum_{k=1}^{2S} [\mathbf{w}_p^{[f]^*}]_{2k} [\boldsymbol{\omega}_p^{[f]}]_{2k} + [\mathbf{g}_p^{[f]^*}]_2 \sum_{k=1}^{2V} [\boldsymbol{\varsigma}_p^{[f]^*}]_{2k} [\boldsymbol{\varrho}_p^{[f]}]_{2k}. \end{aligned} \tag{D.4}$$

Considering (D.4) and $\frac{\partial h^{[f]}(\mathbf{g}^{[f]})}{\partial [\mathbf{g}_p^{[f]}]_1}$ leads to (4.27). Estimation of $[\mathbf{g}_p^{[f]}]_2$ is similar, except that we need to consider the following quantities

- $\mathbf{m}_{pq}^{[f]} = \boldsymbol{\Omega}^{[f]^{-1}} \left(\mathbf{G}_q^{[f]^*} \otimes \mathbf{E}_2\right) \mathbf{w}_{pq}^{[f]}$
- $\tilde{\mathbf{w}}_{qp}^{[f]} = \boldsymbol{\Omega}^{[f]^{-1}} \left(\mathbf{E}_2 \otimes \mathbf{G}_q^{[f]}\right) \mathbf{w}_{qp}^{[f]}$
- $\mathbf{E}_2 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$.

Bibliography

[1] D. Mary, C. Ferrari, and A. Dabbech, "3D reconstruction in the perspective of the SKA," in *ANR MAGELLAN - Kick Off Meeting*, Laboratoire Lagrange, Université de Nice Sophia Antipolis, OCA, CNRS, 2015.

[2] S. Yatawatta and S. Kazemi, "Robust radio interferometric calibration," in *International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, Florence, Italy, 2014, pp. 5392–5396.

[3] C. Lonsdale, "Calibration approaches," LFD memo 015, MIT Haystack, Tech. Rep., 2004.

[4] S. J. Wijnholds, S. van der Tol, R. Nijboer, and A.-J. van der Veen, "Calibration challenges for future radio telescopes," *IEEE Signal Processing Magazine*, vol. 27, no. 1, pp. 30–42, 2010.

[5] J. P. Hamaker, J. D. Bregman, and R. J. Sault, "Understanding radio polarimetry. I. Mathematical foundations," *Astronomy & Astrophysics Supplement Series*, vol. 117, no. 1, pp. 137–147, 1996.

[6] R. Jongerius, S. Wijnholds, R. Nijboer, and H. Corporaal, "An end-to-end computing model for the Square Kilometre Array," *IEEE Computer Society*, vol. 47, no. 9, pp. 48–54, 2014.

[7] M. P. Van Haarlem *et al.*, "LOFAR: The LOw-Frequency ARray," *Astronomy & Astrophysics*, vol. 556, no. A2, 2013.

[8] P. E. Dewdney, P. J. Hall, R. T. Schilizzi, and T. J. L. Lazio, "The Square Kilometre Array," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1482–1496, 2009.

[9] J. E. Noordam, "LOFAR calibration challenges," in *Proceedings of the SPIE: Ground-based Telescopes*, vol. 5489, Glasgow, 2004, pp. 817–825.

[10] S. Bhatnagar, "Calibration and imaging challenges at low radio frequencies: An overview of the state of the art," *The Low Frequency Radio Universe, eds. D.J. Saikia, D. Green, Y. Gupta, & T. Venturi (ASP Conference Series LFRU)*, 2009.

[11] S. J. Wijnholds and A.-J. van der Veen, "Fundamental imaging limits of radio telescope arrays," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 613–623, 2008.

[12] ——, "Multisource self-calibration for sensor arrays," *IEEE Transactions on Signal Processing*, vol. 57, no. 9, pp. 3512–3522, 2009.

[13] A.-J. Boonstra and A.-J. van der Veen, "Gain calibration methods for radio telescope arrays," *IEEE Transactions on Signal Processing*, vol. 51, no. 1, pp. 25–38, 2003.

[14] S. Yatawatta, S. Zaroubi, G. de Bruyn, L. Koopmans, and J. Noordam, "Radio interferometric calibration using the SAGE algorithm," in *13th Digital signal Processing Workshop and 5th Signal Processing Education Workshop*, Marco Island, FL, 2009, pp. 150–155.

[15] J. Raza, A.-J. Boonstra, and A.-J. Van der Veen, "Spatial filtering of RF interference in radio astronomy," *IEEE Signal Processing Letters*, vol. 9, no. 2, pp. 64–67, 2002.

[16] E. Ollila, D. E. Tyler, V. Koivunen, and H. Poor, "Complex elliptically symmetric distributions survey, new results and applications," *IEEE Transactions on Signal Processing*, vol. 60, no. 11, pp. 5597–5625, 2012.

[17] E. Jay, "Détection en environnement non gaussien," Ph.D. dissertation, Université de Cergy Pontoise, 2002.

[18] O. M. Smirnov, "Revisiting the radio interferometer measurement equation. I. A full-sky Jones formalism," *Astronomy & Astrophysics*, vol. 527, no. A106, 2011.

[19] L. M. Ker, "Radio AGN evolution with low frequency radio surveys," Ph.D. dissertation, The University of Edinburgh, 2012.

[20] X. Zhang, M. N. El Korso, and M. Pesavento, "MIMO radar target localization and performance evaluation under SIRP clutter," *Signal Processing Journal*, vol. 130, pp. 217–232, 2017.

[21] S. Van der Tol, "Bayesian estimation for ionospheric calibration in radio astronomy," Ph.D. dissertation, TU Delft, Delft University of Technology, 2009.

[22] E. Conte, A. De Maio, and G. Ricci, "Recursive estimation of the covariance matrix of a compound-gaussian process and its application to adaptive CFAR detection," *IEEE Transactions on Signal Processing*, vol. 50, no. 8, pp. 1908–1915, 2002.

[23] F. Pascal, Y. Chitour, J.-P. Ovarlez, P. Forster, and P. Larzabal, "Covariance structure maximum-likelihood estimates in compound gaussian noise: Existence and algorithm analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 34–48, 2008.

[24] J. E. Noordam, "The measurement equation of a generic radio telescope, AIPS++ implementation note nr 185," ASTRON, Tech. Rep., 1996.

[25] O. M. Smirnov, "Revisiting the radio interferometer measurement equation. II. Calibration and direction-dependent effects," *Astronomy & Astrophysics*, vol. 527, no. A107, 2011.

[26] A. R. Thompson, J. M. Moran, and G. W. Swenson Jr., *Interferometry and synthesis in radio astronomy, 2nd ed.* New York: John Wiley & Sons, 2001.

[27] T. K. Moon, "The Expectation-Maximization algorithm," *IEEE Signal Processing Magazine*, vol. 13, no. 6, pp. 47–60, 1996.

[28] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 36, no. 4, pp. 477–489, 1988.

[29] J. Nocedal and S. J. Wright, *Numerical optimization, 2nd ed.* Springer Science & Business Media, 2006.

[30] J. Friedman, T. Hastie, H. Höfling, R. Tibshirani *et al.*, "Pathwise coordinate optimization," *The Annals of Applied Statistics*, vol. 1, no. 2, pp. 302–332, 2007.

[31] S. Yatawatta, "Reduced ambiguity calibration for LOFAR," *Experimental Astronomy*, vol. 34, no. 1, pp. 89–103, 2012.

[32] S. Kazemi, P. Hurley, O. Öçal, and G. Cherubini, "Blind calibration for radio interferometry using convex optimization," in *3rd International Workshop on Compressed Sensing Theory and its Applications to Radar, Sonar and Remote Sensing (CoSeRa)*, Pisa, Italy, 2015, pp. 164–168.

[33] S. van der Tol and A.-J. van der Veen, "Ionospheric calibration for the LOFAR radio telescope," in *International Symposium on Signals, Circuits and Systems (ISSCS)*, vol. 2, 2007, pp. 1–4.

[34] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and distributed computation: numerical methods*. Prentice Hall, 1989.

[35] S. Yatawatta, "Distributed radio interferometric calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 449, no. 4, pp. 4506–4514, 2015.

[36] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends® in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[37] J. E. Noordam and O. M. Smirnov, "The MeqTrees software system and its use for third-generation calibration of radio interferometers," *Astronomy & Astrophysics*, vol. 524, no. A61, 2010.

[38] B. C. Ng and A. Nehorai, "Active array sensor localization," *Elsevier Signal processing*, vol. 44, no. 3, pp. 309–327, 1995.

[39] J. T. H. Lo and S. L. Marple, "Eigenstructure methods for array sensor localization," in *International Conference on Acoustics, Speech, and Signal Processing (IEEE ICASSP)*, Dallas, Texas, 1987, pp. 2260–2263.

[40] L. Qiong, G. Long, and Y. Zhongfu, "An overview of self-calibration in sensor array processing," in *6th International Symposium on Antennas, Propagation and EM Theory*, Beijing, China, 2003, pp. 279–282.

[41] S. Kazemi and S. Yatawatta, "Robust radio interferometric calibration using the t-distribution," *Monthly Notices of the Royal Astronomical Society*, vol. 435, no. 1, pp. 597–605, 2013.

[42] M. de Vos, A. W. Gunst, and R. Nijboer, "The LOFAR telescope: System architecture and signal processing," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1431–1437, 2009.

[43] P. Zarka *et al.*, "NenUFAR: Instrument description and science case," in *International Conference on Antenna Theory and Techniques (IEEE ICATT)*, Kharkiv, Ukraine, 2015.

[44] C. Ferrari *et al.*, "French SKA White Book - The French community towards the Square Kilometre Array," *Published by the SKA-France Coordination in collaboration with AS SKA-LOFAR, arXiv:1712.06950*, 2017.

[45] R. Levanda and A. Leshem, "Synthetic aperture radio telescopes," *IEEE Signal Processing Magazine*, vol. 27, no. 1, pp. 14–29, 2010.

[46] F. Zernike, "The concept of degree of coherence and its application to optical problems," *Physica*, vol. 5, no. 8, pp. 785–795, 1938.

[47] M. Born and E. Wolf, *Principles of optics - Electromagnetic theory of propagation, interference and diffraction of light - seventh (expanded) edition.* Cambridge University Press, 2000.

[48] S. Meimon, "Reconstruction d'images astronomiques en interférométrie optique," Ph.D. dissertation, Université Paris Sud, 2005.

[49] J. McMullin, B. Waters, D. Schiebel, W. Young, and K. Golap, "CASA architecture and applications," in *Astronomical data analysis software and systems XVI, ASP Conference Series*, vol. 376, 2007, pp. 127–130.

[50] S. Myers, J. Ott, and N. Elias, "CASA synthesis & single dish reduction cookbook," 2010.

[51] N. G. L. Costa Ferreirinha Rocha, "Calibration and Image Reconstruction with Radio Interferometers," Ph.D. dissertation, Université Technique de Lisbonne, 2011.

[52] T. J. Cornwell, K. Golap, and S. Bhatnagar, "The noncoplanar baselines effect in radio interferometry: The W-projection algorithm," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 647–657, 2008.

[53] S. J. Wijnholds, A.-J. van der Veen, F. De Stefani, E. La Rosa, and A. Farina, "Signal processing challenges for radio astronomical arrays," in *International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, Florence, Italy, 2014, pp. 5382–5386.

[54] K. Davies, *Ionospheric radio.* London: Peter Peregrinus, 1990.

[55] H.-S. Jung and W.-J. Lee, "An improvement of ionospheric phase correction by multiple-aperture interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 9, pp. 4952–4960, 2015.

[56] A. E. Kimball and Ž. Ivezić, "A unified catalog of radio objects detected by NVSS, FIRST, WENSS, GB6, and SDSS," *The Astronomical Journal*, vol. 136, no. 2, pp. 684–712, 2008.

[57] J. Baars, R. Genzel, I. Pauliny-Toth, and A. Witzel, "The absolute spectrum of Cas A-an accurate flux density scale and a set of secondary calibrators," *Astronomy & Astrophysics*, vol. 61, no. 1, pp. 99–106, 1977.

[58] A.-J. van der Veen, A. Leshem, and A.-J. Boonstra, "Signal processing for radio astronomical arrays," in *Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM)*, 2004, pp. 1–10.

[59] A.-J. van Der Veen, A. Leshem, and A.-J. Boonstra, "Array signal processing for radio astronomy," in *The Square Kilometre Array: An Engineering Perspective.* Springer, 2005, pp. 231–249.

[60] A.-J. van der Veen and S. J. Wijnholds, "Signal processing tools for radio astronomy," in *Handbook of Signal Processing Systems.* Springer, 2013, pp. 421–463.

[61] S. Salvini and S. J. Wijnholds, "Fast gain calibration in radio astronomy using alternating direction implicit methods: Analysis and applications," *Astronomy & Astrophysics*, vol. 571, no. A97, 2014.

[62] ——, "StEFCal an alternating direction implicit method for fast full polarization array calibration," in *XXXIth URSI General Assembly and Scientific Symposium (URSI GASS)*, Beijing, 2014, pp. 1–4.

[63] U. Rau, S. Bhatnagar, M. A. Voronkov, and T. J. Cornwell, "Advances in calibration and imaging techniques in radio interferometry," *Proceedings of the IEEE*, vol. 97, no. 8, pp. 1472–1481, 2009.

[64] C. Tasse, S. van der Tol, J. van Zwieten, G. van Diepen, and S. Bhatnagar, "Applying full polarization A-Projection to very wide field of view instruments: An imager for LOFAR," *Astronomy & Astrophysics*, vol. 553, no. A105, 2013.

[65] S. Bhatnagar, U. Rau, and K. Golap, "Wide-field wide-band interferometric imaging: The WB A-projection and hybrid algorithms," *The Astrophysical Journal*, vol. 770, no. 2, 2013.

[66] R. Sault and T. Cornwell, "The Hamaker-Bregman-Sault measurement equation," in *Synthesis Imaging in Radio Astronomy II*, vol. 180, 1999, p. 657.

[67] S. Yatawatta, "LOFAR beamshapes and their use in calibration and imaging," ASTRON, Tech. Report, Tech. Rep., 2007.

[68] S. Kazemi, S. Yatawatta, and S. Zaroubi, "Clustered calibration: an improvement to radio interferometric direction-dependent self-calibration," *Monthly Notices of the Royal Astronomical Society*, vol. 430, no. 2, pp. 1457–1472, 2013.

[69] S. Bhatnagar, T. Cornwell, K. Golap, and J. M. Uson, "Correcting direction-dependent gains in the deconvolution of radio interferometric images," *Astronomy & Astrophysics*, vol. 487, no. 1, pp. 419–429, 2008.

[70] M. Lochner, I. Natarajan, J. T. L. Zwart, O. Smirnov, B. A. Bassett, N. Oozeer, and M. Kunz, "Bayesian inference for radio observations," *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 2, pp. 1308–1319, 2015.

[71] J. P. Leahy and R. A. Perley, "VLA images of 23 extragalactic radio sources," *The Astronomical Journal*, vol. 102, pp. 537–561, 1991.

[72] S. Yatawatta, "On the interpolation of calibration solutions obtained in radio interferometry," *Monthly Notices of the Royal Astronomical Society*, vol. 428, no. 1, pp. 828–833, 2013.

[73] S. van der Tol, B. D. Jeffs, and A.-J. van der Veen, "Self-calibration for the LOFAR radio astronomical array," *IEEE Transactions on Signal Processing*, vol. 55, no. 9, pp. 4497–4510, 2007.

[74] R. Nijboer and J. Noordam, "LOFAR calibration," in *Astronomical Data Analysis Software and Systems XVI*, vol. 376, 2007, pp. 237–240.

[75] B. Friedlander and A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE Transactions on Antennas and Propagation*, vol. 39, no. 3, pp. 273–284, 1991.

[76] B. C. Ng and C. M. S. See, "Sensor-array calibration using a maximum-likelihood approach," *IEEE Transactions on Antennas and Propagation*, vol. 44, no. 6, pp. 827–835, 1996.

[77] D. R. Fuhrmann, "Estimation of sensor gain and phase," *IEEE Transactions on Signal Processing*, vol. 42, no. 1, pp. 77–87, 1994.

[78] S. J. Wijnholds and A.-J. Boonstra, "A multisource calibration method for phased array radio telescopes," in *4th Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM)*, 2006, pp. 200–204.

[79] J. E. Noordam, "Generalized self-calibration for LOFAR," *General Assembly of the International Union of Radio Science (URSI), Maastricht,The Netherlands*, 2002.

[80] B. D. Jeffs, S. van der Tol, and A.-J. van der Veen, "Direction dependent self calibration of large distributed sensor arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, Toulouse, France, 2006.

[81] D. A. Mitchell, L. J. Greenhill, R. B. Wayth, R. J. Sault, C. J. Lonsdale, R. J. Cappallo, M. F. Morales, and S. M. Ord, "Real-time calibration of the Murchison Widefield Array," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 707–717, 2008.

[82] S. Bhatnagar, T. Cornwell, and K. Golap, "Solving for the antenna based pointing errors," Technical report, EVLA Memo 84, Tech. Rep., 2004.

[83] A. Cohen, W. Lane, W. Cotton, N. Kassim, T. Lazio, R. Perley, J. Condon, and W. Erickson, "The VLA low-frequency sky survey," *The Astronomical Journal*, vol. 134, no. 3, pp. 1–53, 2007.

[84] W. D. Cotton, J. J. Condon, R. A. Perley, N. Kassim, J. Lazio, A. Cohen, W. Lane, and W. C. Erickson, "Beyond the isoplanatic patch in the VLA low-frequency sky survey," in *SPIE Astronomical Telescopes+ Instrumentation*, vol. 5489. International Society for Optics and Photonics, 2004, pp. 180–189.

[85] W. Cotton and J. Condon, "Calibration and imaging of 74 MHz data from the Very Large Array," in *Proceedings of the URSI General Assembly*, 2002.

[86] H. Intema, S. van der Tol, W. Cotton, A. Cohen, I. van Bemmel, and H. Röttgering, "Ionospheric calibration of low frequency radio interferometric observations using the peeling scheme-i. method description and first results," *Astronomy & Astrophysics*, vol. 501, no. 3, pp. 1185–1205, 2009.

[87] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Processing*, vol. 8, no. 3, pp. 185–210, 1998.

[88] K. Madsen, H. B. Nielsen, and O. Tingleff, "Methods for non-linear least squares problems," Informatics and Mathematical Modelling, Technical University of Denmark, Tech. Rep., 2004.

[89] S. Kazemi, S. Yatawatta, S. Zaroubi, P. Lampropoulos, A. de Bruyn, L. Koopmans, and J. Noordam, "Radio interferometric calibration using the SAGE algorithm," *Monthly Notices of the Royal Astronomical Society*, vol. 414, no. 2, pp. 1656–1666, 2011.

[90] H. Gavin, "The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems," Duke University, 2011.

[91] K. Levenberg, "A method for the solution of certain non–linear problems in least squares," *Quaterly of Applied Mathematics*, vol. 2, pp. 164–168, 1944.

[92] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the Society for Industrial and Applied Mathematics*, vol. 11, no. 2, pp. 431–441, 1963.

[93] S. Yatawatta, S. Kazemi, and S. Zaroubi, "GPU accelerated nonlinear optimization in radio interferometric calibration," in *Innovative Parallel Computing (InPar)*. San Jose, CA: IEEE, 2012, pp. 1–6.

[94] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in *Radar Array Processing*, S. Haykin, J. Litva, and T. J. Shepherd, Eds. Berlin: Springer-Verlag, 1993, ch. 4, pp. 99–151.

[95] S. J. Wijnholds, "Fish-eye observing with phased array radio telescopes," Ph.D. dissertation, Technische Universiteit Delft, Delft, The Netherlands, 2010.

[96] P. E. Gill, W. Murray, and M. H. Wright, "Practical optimization," 1981.

[97] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor array processing using Weighted Subspace Fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436–2449, 1991.

[98] M. Haardt, M. Pesavento, F. Roemer, and M. N. El Korso, "Subspace methods and exploitation of special array structures," in *Electronic Reference in Signal Processing:Array and Statistical Signal Processing (M. Virberg, ed.)*. Academic Press Library in Signal Processing, Elsevier Ltd., 2014, vol. 3.

[99] A. Repetti, J. Birdi, A. Dabbech, and Y. Wiaux, "Non-convex optimization for self-calibration of direction-dependent effects in radio interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 470, no. 4, pp. 3981–4006, 2017.

[100] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.

[101] T. W. Anderson, *An introduction to multivariate statistical analysis*. Wiley New York, 1958, vol. 2.

[102] J. A. Fessler and A. O. Hero, "Space-alternating generalized expectation-maximization algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2664–2677, 1994.

[103] S. Kirkpatrick, "Optimization by simulated annealing: Quantitative studies," *Journal of Statistical Physics*, vol. 34, no. 5, pp. 975–986, 1984.

[104] P.-J. Chung and S. Wan, "Array self-calibration using SAGE algorithm," in *5th Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM)*, Darmstadt, 2008, pp. 165–169.

[105] A. Leshem, A.-J. van der Veen, and A.-J. Boonstra, "Multichannel interference mitigation techniques in radio astronomy," *The Astrophysical Journal Supplement Series*, vol. 131, no. 1, pp. 355–373, 2000.

[106] A. Leshem and A.-J. van der Veen, "Radio-astronomical imaging in the presence of strong radio interference," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1730–1747, 2000.

[107] A.-J. Boonstra, "Radio frequency interference mitigation in radio astronomy," Ph.D. dissertation, Technische Universiteit Delft, Delft, The Netherlands, 2005.

[108] A. M. Zoubir, V. Koivunen, Y. Chakhchoukh, and M. Muma, "Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts," *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.

[109] K. L. Lange, R. J. A. Little, and J. M. G. Taylor, "Robust statistical modeling using the t distribution," *Journal of the American Statistical Association*, vol. 84, no. 408, pp. 881–896, 1989.

[110] C. Liu and D. B. Rubin, "ML estimation of the t distribution using EM and its extensions, ECM and ECME," *Statistica Sinica*, vol. 5, no. 1, pp. 19–39, 1995.

[111] T. L. Grobler, C. Nunhokee, O. Smirnov, A. Van Zyl, and A. De Bruyn, "Calibration artefacts in radio interferometry– I. Ghost sources in Westerbork Synthesis Radio Telescope data," *Monthly Notices of the Royal Astronomical Society*, vol. 439, no. 4, pp. 4030–4047, 2014.

[112] A. Bartkowiak, "Should normal distribution be normal? The Student's t alternative," in *6th International Conference on Computer Information Systems and Industrial Management Applications (CISIM)*, Minneapolis, MN, 2007, pp. 3–8.

[113] L. Li, X. Wang, and G. Wang, "Alternating Direction Method of Multipliers for separable convex optimization of real functions in complex variables," *Mathematical Problems in Engineering*, vol. 2015, no. Article ID 104531, pp. 1–14, 2015.

[114] B. Jiang, S. Ma, and S. Zhang, "Alternating direction method of multipliers for real and complex polynomial optimization models," *Optimization*, vol. 63, no. 6, pp. 883–898, 2014.

[115] P.-A. Absil, C. G. Baker, and K. A. Gallivan, "Trust-region methods on Riemannian manifolds," *Foundations of Computational Mathematics*, vol. 7, no. 3, pp. 303–330, 2007.

[116] R. E. Carrillo, J. D. McEwen, and Y. Wiaux, "PURIFY: a new approach to radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 439, no. 4, pp. 3591–3604, 2014.

[117] A. Ferrari, D. Mary, R. Flamary, and C. Richard, "Distributed image reconstruction for very large arrays in radio astronomy," in *8th Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM)*, 2014, pp. 389–392.

[118] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy & Astrophysics Supplement Series*, vol. 15, pp. 417–426, 1974.

[119] T. Cornwell and K. Evans, "A simple maximum entropy deconvolution algorithm," *Astronomy & Astrophysics*, vol. 143, pp. 77–83, 1985.

[120] C. Ben-David and A. Leshem, "Parametric high resolution techniques for radio astronomical imaging," *IEEE Journal of Selected Topics in Signal Processing*, vol. 2, no. 5, pp. 670–684, 2008.

[121] R. Levanda and A. Leshem, "Radio astronomical image formation using sparse reconstruction techniques," in *IEEE 25th Convention of Electrical and Electronics Engineers in Israel*, 2008, pp. 716–720.

[122] R. Carrillo, J. McEwen, and Y. Wiaux, "Sparsity Averaging Reweighted Analysis (SARA): a novel algorithm for radio-interferometric imaging," *Monthly Notices of the Royal Astronomical Society*, vol. 426, no. 2, pp. 1223–1234, 2012.

[123] A. Dabbech, C. Ferrari, D. Mary, E. Slezak, O. Smirnov, and J. S. Kenyon, "MORESANE: MOdel REconstruction by Synthesis-ANalysis Estimators-a sparse deconvolution algorithm for radio interferometric imaging," *Astronomy & Astrophysics*, vol. 576, no. A7, 2015.

[124] U. Rau and T. J. Cornwell, "A multi-scale multi-frequency deconvolution algorithm for synthesis imaging in radio interferometry," *Astronomy & Astrophysics*, vol. 532, no. A71, 2011.

[125] G. Loose, "LOFAR Self-Calibration using a Blackboard software architecture," in *Astronomical Data Analysis Software and Systems XVII ASP Conference Series*, vol. 394, 2008, pp. 91–94.

[126] V. Pandey, J. van Zwieten, A. de Bruyn, and R. Nijboer, "Calibrating LOFAR using the Black Board Selfcal System," in *The Low-Frequency Radio Universe*, vol. 407, 2009, pp. 384–388.

[127] G. J. McLachlan and T. Krishnan, *The EM algorithm and extensions, 2nd ed.* John Wiley & Sons, 2008.

[128] A. Offringa, A. De Bruyn, M. Biehl, S. Zaroubi, G. Bernardi, and V. Pandey, "Post-correlation radio frequency interference classification methods," *Monthly Notices of the Royal Astronomical Society*, vol. 405, no. 1, pp. 155–167, 2010.

[129] I. Martí-Vidal and J. Marcaide, "Spurious source generation in mapping from noisy phase-self-calibrated data," *Astronomy & Astrophysics*, vol. 480, no. 1, pp. 289–295, 2008.

[130] B. Meriaux, C. Ren, M. N. El Korso, A. Breloy, and P. Forster, "Robust-COMET for covariance estimation in convex structures: Algorithm and statistical properties," in *International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (IEEE CAMSAP)*, Curaçao, Dutch Antilles, 2017.

[131] ——, "Efficient estimation of scatter matrix with convex structure under t-distribution," in *International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, Calgary, Canada, 2018.

[132] K. Yao, "Spherically invariant random processes: Theory and applications," in *Communications, Information and Network Security*, V. Bhargava *et al.*, Eds. Dordrecht, The Netherlands: Kluwer Academic: Springer, 2003, pp. 315–331.

[133] J. Wang, A. Dogandzic, and A. Nehorai, "Maximum likelihood estimation of compound-gaussian clutter and target parameters," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3884–3898, 2006.

[134] C. D. Nunhokee, "Link between ghost artefacts, source suppression and incomplete calibration sky models," Ph.D. dissertation, Rhodes University, Rhodes, South Africa, 2015.

[135] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Joint ML calibration and DOA estimation with separated arrays," in *International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, Shanghai, China, 2016.

[136] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, P. Larzabal, and M. Pesavento, "Bayesian calibration using different prior distributions: an iterative maximum a posteriori approach for radio interferometers," in *26th European Signal Processing Conference (EUSIPCO)*, Rome, Italy, 2018.

[137] V. Ollier, R. Boyer, M. N. El Korso, and P. Larzabal, "Bayesian lower bounds for dense or sparse (oulier) noise in the RMT framework," in *9th Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM)*, Rio de Janeiro, Brazil, 2016.

[138] V. Ollier, M. N. El Korso, R. Boyer, P. Larzabal, and M. Pesavento, "Relaxed concentrated MLE for robust calibration of radio interferometers," in *24th European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, 2016, pp. 280–284.

[139] ——, "Robust calibration of radio interferometers in non-Gaussian environment," *IEEE Transactions on Signal Processing*, vol. 65, no. 21, pp. 5649–5660, 2017.

[140] A. Hjorungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Transactions on Signal Processing*, vol. 55, no. 6, pp. 2740–2746, 2007.

[141] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Processing Magazine*, vol. 33, no. 1, pp. 57–77, 2016.

[142] D. P. Bertsekas, *Nonlinear programming*. Athena scientific Belmont, 1999.

[143] S. Vorobyov, A. B. Gershman, and K. M. Wong, "Maximum likelihood direction-of-arrival estimation in unknown noise fields using sparse sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 34–43, 2005.

[144] A. Balleri, A. Nehorai, and J. Wang, "Maximum likelihood estimation for compound-Gaussian clutter with inverse gamma texture," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 2, pp. 775–779, 2007.

[145] P. Stoica and R. L. Moses, *Spectral analysis of signals*. Pearson Prentice Hall, Upper Saddle River, NJ, 2005.

[146] O. Besson and Y. I. Abramovich, "On the Fisher information matrix for multivariate elliptically contoured distributions," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1130–1133, 2013.

[147] M. N. El Korso, A. Renaux, and P. Forster, "CRLB under K-distributed observation with parameterized mean," in *8th Sensor Array and Multichannel Signal Processing Workshop (IEEE SAM)*, A Coruña, Spain, 2014.

[148] S. Li, H. Wang, and T. Chai, "A t-distribution based particle filter for target tracking," in *American Control Conference*, Minneapolis, MN, 2006, pp. 2191–2196.

[149] M. Brossard, M. N. El Korso, M. Pesavento, R. Boyer, P. Larzabal, and S. J. Wijnholds, "Parallel multi-wavelength calibration algorithm for radio astronomical arrays," *Elsevier Signal Processing Journal*, vol. 145, pp. 258–271, 2018.

[150] C. M. Trott and R. B. Wayth, "Spectral calibration requirements of radio interferometers for Epoch of Reionisation science with the SKA," *Publications of the Astronomical Society of Australia*, vol. 33, no. e019, 2016.

[151] V. Ollier, M. N. El Korso, A. Ferrari, R. Boyer, and P. Larzabal, "Robust distributed calibration of radio interferometers with direction dependent distortions," *accepted for publication by Elsevier Signal Processing*, 2018.

[152] ——, "Robust calibration of radio interferometers in multi-frequency scenario," in *International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, Calgary, Canada, 2018.

[153] S. Yatawatta, "Fine tuning consensus optimization for distributed radio interferometric calibration," in *24th European Signal Processing Conference (EUSIPCO)*, Budapest, Hungary, 2016.

[154] C. Bilen, G. Puy, R. Gribonval, and L. Daudet, "Convex optimization approaches for blind sensor calibration using sparsity," *IEEE Transactions on Signal Processing*, vol. 62, no. 18, pp. 4847–4856, 2014.

[155] M. Brossard, M. N. El Korso, M. Pesavento, R. Boyer, and P. Larzabal, "Calibration of radio interferometers using a sparse doa estimation framework," in *24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 275–279.

[156] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[157] S. Yatawatta, "Radio interferometric calibration using a riemannian manifold," in *International Conference on Acoustics, Speech and Signal Processing (IEEE ICASSP)*, 2013, pp. 3866–3870.

[158] S. M. Kay, *Fundamentals of statistical signal processing: estimation theory*. Prentice Hall, 1993.

[159] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 451–462, 2009.

[160] F. Pukelsheim, *Optimal design of experiments*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 1993.

[161] J. Li, L. Xu, P. Stoica, K. W. Forsythe, and D. W. Bliss, "Range compression and waveform optimization for mimo radar: a cramer–rao bound based study," *IEEE Transactions on Signal Processing*, vol. 56, no. 1, pp. 218–232, 2008.

[162] A. Mennad, S. Fortunati, M. N. El Korso, A. Younsi, A. M. Zoubir, and A. Renaux, "Slepian-bangs-type formulas and the related misspecified Cramér-Rao Bounds for Complex Elliptically Symmetric distributions," *Elsevier Signal Processing*, vol. 142, pp. 320–329, 2018.

Titre : Contributions aux méthodes de calibration robuste en radioastronomie
Mots clés : Estimation, calibration robuste, optimisation distribuée, radioastronomie

Résumé : En radioastronomie, les signaux d'intérêt mesurés par les interféromètres sont perturbés par de nombreux effets environnementaux et instrumentaux, nécessitant la mise en oeuvre de techniques algorithmiques pour les traiter et pouvoir ainsi reconstruire *in fine* des images parfaitement nettes de l'espace. Cette étape de correction des perturbations se nomme la calibration et repose généralement sur une modélisation gaussienne du bruit, pour une seule fréquence considérée. Cependant, en pratique, cette l'hypothèse n'est pas toujours valide car de multiples sources inconnues de faible intensité sont visibles dans le champ de vision et des interférences radioélectriques perturbent les données. En outre, réaliser une calibration indépendante, fréquence par fréquence, n'est pas la manière la plus optimale de procéder.

Le but de ce travail est donc de développer des algorithmes de correction dans le traitement des signaux radio qui soient robustes à la présence d'éventuelles valeurs aberrantes ou sources d'interférences, et qui soient adaptés au contexte multi-fréquentiel.
Par conséquent, nous nous appuyons sur une modélisation plus générale que la loi gaussienne, appelée processus Gaussien composé, et proposons un algorithme itératif basé sur l'estimation au sens du maximum de vraisemblance. En accord avec le scénario multi-fréquentiel sous étude, nous exploitons la variation spectrale des perturbations en utilisant des méthodologies telles que l'optimisation distribuée sous contraintes et le traitement parallèle des données.

Title: Contributions to robust calibration methods in radio astronomy
Keywords: Estimation, robust calibration, distributed optimization, radio astronomy

Abstract: Accurate calibration is of critical importance for new advanced interferometric systems in radio astronomy in order to recover high resolution images with no distortions. This process consists in correcting for all environmental and instrumental effects which corrupt the observations. Most state-of-the-art calibration approaches assume a Gaussian noise model and operate mostly in an iterative manner for a mono-frequency scenario. However, in practice, the Gaussian classical noise assumption is not valid as radio frequency interference affects the measurements and multiple unknown weak sources appear within the wide field-of-view. Furthermore, considering one frequency bin at a time with a single centralized agent processing all data leads to suboptimality and computational limitations.

The goal of this thesis is to explore robustness of calibration algorithms w.r.t. the presence of outliers in a multi-frequency scenario. To this end, we propose the use of an appropriate noise model, namely, the so-called compound-Gaussian which encompasses a broad range of different heavy-tailed distributions. To combine limited computational complexity and quality of calibration, we designed an iterative calibration algorithm based on the maximum likelihood estimator under the compound-Gaussian modeling. In addition, a computationally efficient way to handle multiple sub-frequency bands is to apply distributed and decentralized strategies. Thus, the global operational load is distributed over a network of computational agents and calibration amounts to solve a global constrained problem thanks to available variation models or by assuming smoothness across frequency.

Université Paris-Saclay
Espace Technologique / Immeuble Discovery
Route de l'Orme aux Merisiers RD 128 / 91190 Saint-Aubin, France