Bayesian Estimation for Ionospheric Calibration in Radio Astronomy

Bayesian Estimation for Ionospheric Calibration in Radio Astronomy

PROEFSCHRIFT

ter verkrijging van de graad van doctor
aan de Technische Universiteit Delft,
op gezag van de Rector Magnificus Prof. dr. ir. J.T. Fokkema,
voorzitter van het College voor Promoties,
in het openbaar te verdedigen op 10 november 2009 om 10.00 uur,
door

Sebastiaan VAN DER TOL

elektrotechnisch ingenieur
geboren te Delft.

Dit proefschrift is goedgekeurd door de promotoren:

Prof. dr. ir. A.J. van der Veen
Prof. dr. H.J.A. Röttgering

Samenstelling promotiecommissie:

Rector Magnificus	voorzitter
Prof. dr. ir. A.J. van der Veen	Technische Universiteit Delft, promotor
Prof. dr. H.J.A. Röttgering	Universiteit Leiden, promotor
Prof. dr. H.W.J. Russchenberg	Technische Universiteit Delft
Prof. dr. ir. C.P.A. Wapenaar	Technische Universiteit Delft
Prof. dr. W.N. Brouw	Rijks Universiteit Groningen
Prof. dr. A. Leshem	Bar-Ilan University
Dr. ir. H van. der Marel	Technische Universiteit Delft

Author email: `vdtol@strw.leidenuniv.nl`

To Kimberley

Contents

i

Introduction

Astronomy is an ancient science that is still thriving today. Even though the objects of study are remote it has had a direct influence on the life of people as it led to the development of calendars and aided navigation. The discovery of laws of motion of the celestial bodies by Copernicus, Kepler and Galilei culminated in the formulation of the laws of classical mechanics by Newton. From then on the knowledge of the laws of nature increased rapidly leading to the technological age.

Apart from its practical consequences, astronomy has given us an understanding of our origin and place in the universe. Even today many new and exciting discoveries are done, ranging from the earliest galaxies at the edge of visible universe, to planets orbiting stars near us or water on our neighbour planet Mars.

All major discoveries in astronomy followed after a development in observing techniques. A great example is the invention of the optical telescope which Galileo Galilei used to observe Jupiter and discover its moons. Nowadays observations are not limited to visible light part of the spectrum. Observations range from radio, infrared, X-ray to gamma radiation. Each frequency band provides a unique view on the universe. Some of these observations are carried out in space out of necessity; for example X-ray and gamma radiation are blocked by the earth atmosphere. Other instruments, such as the Hubble space telescope, are located in space because without the earth atmosphere a much higher image quality can be obtained. An alternative solution is to use adaptive optics for earth based optical telescopes, where the shape of the reflector is adjusted to compensate for the atmospheric distortions.

The very first radio astronomical observations by Karl G. Jansky in 1931 were done at 20.5 MHz with a single mechanically steerable antenna array. Since then radio astronomy has developed tremendously. Nowadays most radiotelescopes are interferometers: they consist of multiple receivers and their signals are correlated with each other. Using a technique called *aperture synthesis* a resolution can be obtained corresponding to that of a single dish of the diameter of the longest baseline. The baselines are the separations between pairs of receivers. In Very Long Baseline Interometry (VLBI) the signals from

receivers across continents are combined, and resolutions of microarcseconds have been obtained. The invention of self calibration made it possible to make images with a dynamic range, i.e. the ratio of the brightest source in the image and the weakest visible detail, of 10^5 [1, page 28].

The observing frequencies currently used are much higher than in Jansky's experiments, ranging from several hundred MHz up to several hundred GHz. A clear advantage of observing at higher frequencies is that the resolution of the images is higher for a given instrument size. But the most important factor that has hampered observations below about 150 MHz is the ionosphere.

The ionosphere is the upper part of the earth atmosphere where radiation from the sun is so intense that it partly ionizes the air. The free electrons change the phase of electromagnetic waves traveling through the ionosphere. As the frequency goes down the phase is affected more and more until the plasma frequency is reached and the wave is completely reflected. Below the plasma frequency earth based observations are impossible. But even long before the plasma frequency is reached observations are already getting increasingly more difficult. Dynamic processes in the ionosphere cause the electron density to vary both in time and over space. Furthermore, the field of view of a radio telescope scales with the wavelength, so at low frequencies the field of view is larger for a fixed diameter of the receivers. And at low frequencies baselines need to be longer to achieve the same resolution. These three effects, larger ionospheric phase changes, larger field of view and larger baselines, taken together cause the ionospheric phase to vary from antenna to antenna, over the field of view and over time. This explains why ionospheric calibration is so much more difficult at low frequencies and hence why it is difficult to make high dynamic range images.

This problem has kept astronomers away from observing at low frequencies for a long time. However a lot of interesting science can be done in these scarcely studied frequency bands. For example the telltale signature of the formation of the first stars starting the Epoch of Reionizaton (EOR) is believed to be found in this band. The increase in processing capacity of digital hardware makes it nowadays possible to overcome the ionospheric calibration problem. A number of efforts are currently underway to build instruments that observe at low frequencies in the range of about 30MHz-250MHz. However just an increase of processing capacity is not sufficient to solve the ionospheric calibration problem. The algorithms still need to be developed. The development of an ionospheric calibration algorithm is the subject of this thesis.

1.1   Current developments

Because of scientific interest in low frequency observation and the increased processing capabilities that make these observations possible a number of instruments for this frequency range has been developed in recent years or is currently being developed. Existing instruments with low frequency capability include

- **MERLIN** Multielement Radio-linked Interferometer Network, Great Britain. This network of 6 radio telescopes is operational since 1980. The lowest frequency band

is at 151 Mhz.

- **GMRT** Giant Metrewave Radio Telescope, India. This telescope is operational since 1999. The lowest available frequency band is at 151 MHz. A 50 MHz extension is being designed.

- **VLA** Very Large Array in New Mexico, U.S.A. - Since 1998 all 27 antennas of the VLA are equiped with a 74 MHz system

- **WSRT** Westerbork Synthesis Radio Telescope in The Netherlands. Since 2004 the Low Frequency Front Ends (LFFE) are operational. These receivers can be tuned in the frequency range 115 MHz - 180 MHz.

Instruments currently under development or construction include

- **LOFAR** Low Frequency Array, 20MHz - 250MHz, The Netherlands. Currently under construction,

- **MWA** Murchison Widefield Array 80MHz-300MHz, 500 (tiles) x 16 (dual polarization dipoles) = 8000 dual-polarization dipole antennas placed in the outback of Western Australia,

- **LWA** Long Wavelength Array, South West U.S.A. 10-88 MHz,

- **Past/21CMA** China. An array specifically designed to search for the EOR signal.

1.2   Ionospheric Calibration

The technique of aperture synthesis makes it possible to obtain the resolution corresponding to an instrument of the size of the largest baseline. The synthesized aperture is only sampled and not completely filled as would be the case for a dish of the size of the array. This results in a point spread function (PSF) that has a high sidelobe level. This limits the dynamic range of the so called dirty image. Much better images can be made by deconvolving the dirty image. This can be done because the PSF is known, at least as long as the array is well calibrated. Successful deconvolution requires an accurate calibration.

In the early days of radio astronomy calibration was done by first pointing the array at a bright calibrator source, observe it for some time and then point the array to the field of interest. The accuracy of this method is limited by the fact that the correction in the direction of the calibrator can be different from the correction needed for the target field. In the early 1980s self calibration methods have been developed [2, 3]. The idea is to use the observations of the target for calibration. Since imaging can be considered as estimating the value of all the pixels in the image, the extension of this problem by estimating one additional complex gain per antenna adds only a few extra degrees of freedom. Self calibration turned out to be enormously successful, increasing the dynamic range by two orders of magnitude, from a few thousand to one, to a few hundred thousand to one.

The field of view of the LOFAR stations is relatively large. The gain is direction dependent and varies rapidly over time. This complicates the calibration problem considerably. In extreme cases the problem becomes untractable.

1.3   Problem Statement and Research Objectives

To satisfy the scientific goals for LOFAR it is essential that high dynamic range images can be made. A limiting factor for the dynamic range are the distortions of the signals by the ionosphere. This thesis is concerned about the following question *How to calibrate the ionospheric phase for LOFAR and other low frequency instruments to obtain high dynamic range images?* Here the scope will be limited to calibration at central level, i.e. a station will be treated as a single directional receiving element. Calibration on station level, where the gains of the antennas need to be determined, is the topic of the thesis by Stefan Wijnholds, which will be published concurrently with this work.

The approach to solve this question is to apply techniques used in signal processing. The usual approach in solving a signal processing problem consists of several steps. These include:

- Making a data model describing the available data in a mathmatical form;

- Formulating an optimality criterion that describes the quality of the estimates of the parameters of interest;

- Finding the best theoretically obtainable performance according to the optimality criterion defined in the previous step. This is the benchmark, or lower bound against which algorithms are evaluated;

- Deriving an algorithm that attains the lower bound. Usually it is not possible to find an algorithm that exactly reaches the bound or the computational costs are excessive. Using approximations one searches for an algorithm that performes close to the bound at an acceptable computational cost;

- Application of the algorithm on simulated data. An analytical proof of the effectiveness of an algorithm can seldom be given. Therefore the algorithm is tested via Monte Carlo runs on simulated data;

- Application of the algorithm on observed data. Since real data never exactly follows the data model the only real proof of the effectiveness of the algorithm can only be obtained by an application to real observed data.

This procedure leads to the following research objectives.

- Find a model for the ionospheric phase fluctuations, detailed enough to model the phases accurately, simple enough to allow the derivation of an estimation algorithm;

- Derive an estimation algorithm for the ionospheric phase that optimizes the dynamic range of the final image;

- Demonstrate the algorithm on simulated data;

- Demonstrate the algorithm on observed data.

1.4   Contributions

The main contribution of this work to the field of ionospheric calibration is the introduction of a statistical model of the ionosphere, not only to generate simulation data, but also to actually derive optimal algorithms to estimate the ionospheric phase.

This thesis describes a Bayesian method of estimating the ionospheric phase fluctuations based on a statistical model of the ionosphere. Observations supporting this model are presented. The method is shown to outperform an existing method based on Zernike polynomials, both on simulated data and observed data. Furthermore it is shown that the method can be extended to more complicated models in a straightforward manner.

1.5   Context

This work was carried out for the VICI-SPCOM project "Signal processing for future wireless communications" [1] within the Circuits and Systems group of the faculty of Electrical Engineering, Mathematics and Computer Science at the Delft University of Technology.

Although radio astronomers share the spectrum with communication applications, the domains of radio astronomy and communication are usually considered as two separate worlds. However, previous cooperation between Circuits and Systems group and Astron, the Netherland Institute for Astronomy in the STW project "Nulling obstructing electromagnetic interferers (NOEMI)" [2] has shown that the signal processing algorithms used for communications can be applied to radio astronomical observations as well.

1.6   List of publications

The following publications were prepared in the context of this thesis.

Journal papers

- S. van der Tol and A.-J. van der Veen. Performance analysis of spatial filtering of RF interference in radio astronomy. *Signal Processing, IEEE Transactions on*, 53 (3):896–910, March 2005. ISSN 1053-587X. doi: 10.1109/TSP.2004.842177.

---

[1] Supported by STW under contract number DTC.5893
[2] Supported by STW under contract number DEL77-476

- S. van der Tol, B.D. Jeffs, and A.-J.. van der Veen. Self-Calibration for the LOFAR Radio Astronomical Array. *Signal Processing, IEEE Transactions on*, 55(9):4497–4510, Sept. 2007. ISSN 1053-587X. doi: 10.1109/TSP.2007.896243.

- C. van der Tol, S. van der Tol, A. Verhoef, B. Su, J. Timmermans, C. Houldcroft, and A. Gieske. A Bayesian approach to estimate sensible and latent heat over vegetation. *Hydrology and Earth System Sciences Discussions*, 6:2337–2365, March 2009.

- H. T. Intema, S. van der Tol, W. D. Cotton, A. S. Cohen, I. M. van Bemmel, and H. J. A. Rottgering. Ionospheric Calibration of Low Frequency Radio Interferometric Observations using the Peeling Scheme: I. Method Description and First Results. *ArXiv e-prints*, April 2009.

- S. van der Tol, R. Sridharan, A.J. van der Veen, H.J.A. Röttgering, and A. S. Cohen. Vlss paper. 2009. In preparation.

Conferences

- A. J. Boonstra, S. J. Wijnholds, S. van der Tol and B. Jeffs. Calibration, Sensitivity and RFI Mitigation Requirements for LOFAR. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2005.

- A.J. Boonstra and S. van der Tol. Spatial Filtering of Interfering Signals at the Initial Low Frequency Array (LOFAR) Phased Array Test Station. In *Radio Science*, volume 40, 2005.

- S. van der Tol and A.J. van der Veen. Application of Robust Capon Beamforming to radio astronomical imaging. In *Proc. IEEE ICASSP*, pages IV–1089–1092, Philadelphia (PA), March 2005. IEEE.

- S. van der Tol, B. Jeffs, and A.J. van der Veen. Calibration of a large distributed low frequency radio astronomical array (LOFAR). In *EUSIPCO*, Antalya (T), sep 2005. Eurasip.

- B. Jeffs, S. van der Tol, and A.J. van der Veen. Direction dependent self calibration of large distributed sensor arrays. In *IEEE ICASSP*, Toulouse (FR), May 2006.

- S. van der Tol and S.J. Wijnholds. CRB Analysis of the Impact of Unknown Receiver Noise on Phased Array Calibration. pages 185–189, July 2006. doi: 10.1109/SAM.2006.1706118.

- S. van der Tol and A.-J. van der Veen. Ionospheric Calibration for the LOFAR Radio Telescope. volume 2, pages 1–4, Iasi, Romania, July 2007. doi: 10.1109/ISSCS.2007.4292761.

1.7   Thesis Outline

Most chapters in this thesis were originally written as self-contained papers. If applicable a reference to the paper and its status (in preparation, accepted or published) is given at the beginning of a chapter. Only minor modifications have been made to include the papers in this thesis, so some overlap between the chapters is still present. Chapter 2 gives an overview of the LOFAR instrument, the ionosphere and calibration in radio astronomy. Chapter 3 describes the "peeling" algorithm which is a sequential least squares estimation algorithm and the current leading candidate algorithm for calibration of LOFAR. The performance of this algorithm is analysed. Depending on the choice for the underlying model for the ionosphere "peeling" can reach satisfactory results. Chapter 4 presents a stochastic model of the ionospheric phase fluctuations and some measurements to validate the model and estimate its key parameters. In Chapter 5 the optimal Bayesian estimator for this model is derived and applied to simulated data. Chapter 6 presents a test setup where peeling is combined with the Bayesian estimator to be applied to 74MHz data of the VLA. Chapter 7 presents possibilities to extend the estimator to a 3D model and a temporal model of the ionosphere. Chapter 8 contains the conclusions and recommendations.

Chapter 2

Preliminaries

This chapter introduces a number of key concepts related to the calibration problem considered in this thesis. First we describe the LOFAR instrument. Then we describe the ionosphere and its effect on radio astronomical observations. Self calibration techniques are described and "Peeling" , a sequential self calibration method for direction dependent calibration is introduced.

2.1 The LOFAR radio telescope

LOFAR is a low frequency radio astronomical array currently under development in The Netherlands by a consortium led by ASTRON. Construction is expected to be finished in 2010. It is designed to produce synthesis images of the most distant (and thus youngest) celestial objects yet observed. LOFAR will observe at unusually low frequencies (15-240 MHz). The LOFAR design calls for an instrument consisting of nearly 13,000 relatively wide field of view small antennas mounted at ground level. These are grouped into stations consisting of 96 dual polarization low band antennas (15-90MHz) and 48 tiles of 16 dual polarization high band antennas (110-240MHz). There will be 18 core stations, 18 remote stations and 8 international stations The core stations packed into a 2-3km area. The remote stations are distributed over the Netherlands. The furthest remote station will be located at aproximately 80km from the core. The international stations are located in Germany (5), Sweden (1), France (1), Great Brittain (1). The maximum distance from the core to the international stations is about 800km

The antennas in each station are used as a phased array and are combined in such a way that a beam is formed into a desired look direction. The resulting output of each beamformer is similar to the output of a telescope dish pointing into the same direction, but is obtained without the use of any moving parts. The beamformer outputs of each station are transported over optical fibers to a central location, where (similar to existing synthesis telescopes [4–7]) they are correlated to the outputs of the other stations, and

9

processed into an image.

2.2   The Ionosphere

The ionosphere is the outer layer of the earth's atmosphere. Radiation from the sun partly ionizes the atmosphere and the resulting free electrons slow down electromagnetic waves propagating through the ionosphere. This additional propagation delay is proportional to the wavelength squared, hence the corresponding phase shift is proportional to the wavelength. At lower frequencies the effects of the ionosphere are more severe. Turbulence in the ionosphere causes the electron density to fluctuate both over time (order 10 seconds) and space (order 10 kilometers).

2.2.1   Electron density and refractive index

The refractive index of a plasma at frequency f is given by [see 8, chap. 8]

$$n = \sqrt{1 - \frac{f_p^2}{f^2}} \tag{2.1}$$

where f_p is the plasma frequency. The plasma frequency squared is proportional to the electron density N_e

$$f_p^2 = \frac{N_e e^2}{4\pi^2 m_e \epsilon_0}, \tag{2.2}$$

where e is the electron charge, m_e is the mass of an electron and ϵ_0 the permittivity of vacuum. When the plasma frequency is much smaller than the frequency of the electromagnetic wave then the deviation of the refractive index from unity can be approximated by

$$\Delta n = n - 1 \approx -\frac{f_p^2}{2f^2} \propto N_e. \tag{2.3}$$

Thus when $f_p \ll f$ the refractive index can assumed to be proportional to the electron density. The change in refractive index causes a propagation delay which leads to a phase rotation given by

$$\phi = \frac{2\pi f}{c} \int (n(\mathbf{s}) - 1)ds = \frac{e^2}{2\pi f c m_e \epsilon_0} \int N_e(\mathbf{s})ds, \tag{2.4}$$

where the integral is along the linge of sight and c is the speed of light in vacuum. The ionospheric phase is approximately proportional to the integrated electron density and inversely proportional to the frequency. The integral in (2.4) is usually refered to as the Total Electron Content (TEC), defined as

$$\text{TEC} = \int N_e(\mathbf{s})ds, \tag{2.5}$$

when the integral is along the vertical axis. For other directions the integral is refered to as the Slant TEC or STEC.

Station beam field of view V

Ionospheric irregularity scale S

Regime 1
V << S
A << S

Array aperture A

(a) Regime 1

Station beam field of view V

Ionospheric irregularity scale S

Regime 2
V << S
A >> S

Array aperture A

(b) Regime 2

Station beam field of view V

Ionospheric irregularity scale S

Regime 3
V >> S
A << S

Array aperture A

(c) Regime 3

Station beam field of view V

Ionospheric irregularity scale S

Regime 4
V >> S
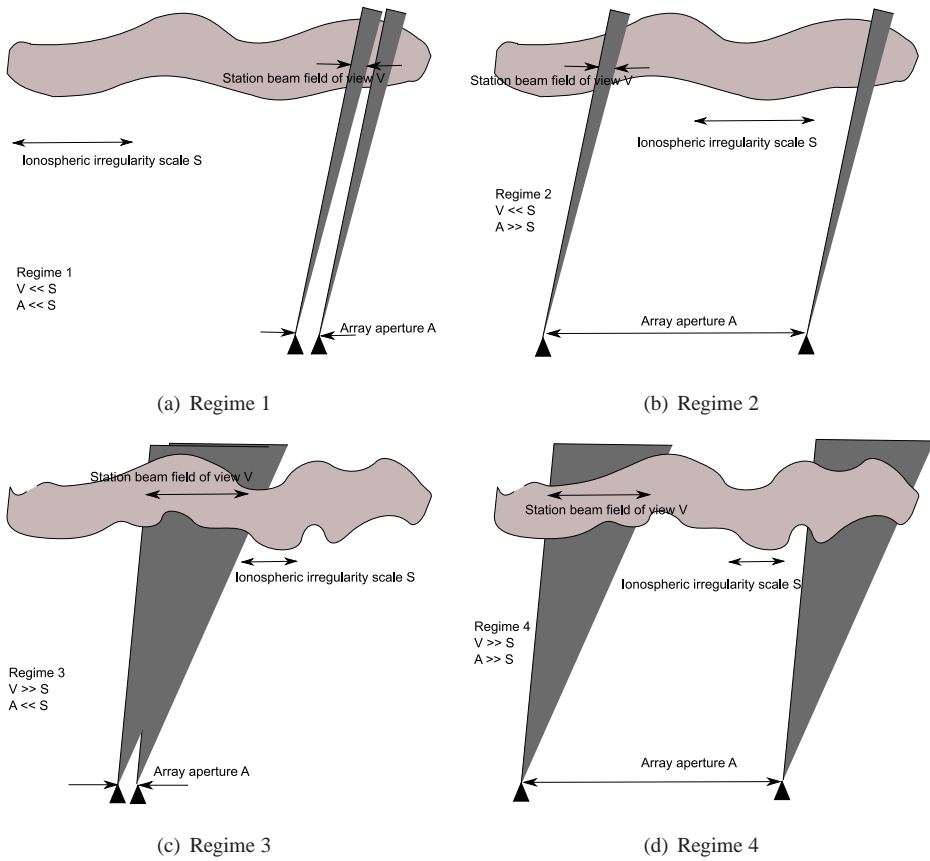A >> S

Array aperture A

(d) Regime 4

Figure 2.1: *Regime 4. The problem of LOFAR calibration through ionospheric refraction. Unknown complex gains through the ionosphere are different for each source at each station (after C. Lonsdale [9]).*

2.2.2 Ionospheric Regimes

The effect of the ionosphere on interferometric observations depends on the relative size of three different measures:

1. the distance between the antennas, or stations in the LOFAR case,

2. the size of the field of view projected onto the ionospheric layer,

3. the scale size of ionospheric fluctuations, that is the distance over which the ionosphere changes significantly.

Based on these three parameters, Lonsdale [9] distinguishes four different regimes. For regime 1, see Figure 2.1(a), both the size of the array and the size of the beam at the ionospheric layer are much smaller than the scale of the ionospheric irregularities. All antennas see the same ionosphere. The ionosphere adds a common phase to all antennes, for all sources. As an interferometer only measure phase differences, this is invisible to an interferometer. In regime 2, see Figure 2.1(b), the spacing between the antennas is larger than the ionospheric scale, but the beam size is smaller. Each antenna sees a different ionosphere. Over the field of view however the ionospheric phase is constant, leading to one unknown phase correction per antenna. In regime 3, see Figure 2.1(c), the beam size is larger than the ionospheric scale, but the antenna spacing is much smaller. The ionospheric phase varies over the field of view, but this ionospheric phase cancels in the correlator. Figure 2.1(d) illustrates how ionospheric phase and gain perturbations affect LOFAR calibration. The ionospheric irregularity scale is smaller than both the full array aperture and individual station beam field of views, so every station and source direction requires a unique calibration solution. This is the regime studied in this thesis.

2.3 Self Calibration Methods

In the radio astronomy literature, "self calibration" (or Selfcal) refers to the calibration of a telescope array using existing sky signals as reference sources [4, 6]. It is assumed that these sources have known position and are relatively bright. The parameters to estimate are the direction-independent electronic gains and phases. Techniques for this have been proposed and are widely used for higher frequency synthesis arrays [2, 10–12], and the estimation statistics are well understood [13]. As a refinement, Selfcal is often combined with the well-known CLEAN algorithm for deconvolution [14, 15], i.e., a technique to iteratively estimate the location of the sources and their powers.

In the array signal processing literature, "self calibration" (or auto-calibration) refers to a much wider class of algorithms, namely calibration using non-cooperative sources. Typically, the location of the sources is considered unknown (cf. Direction of Arrival [DOA] estimation). The additional parameters which need to be estimated can include the direction-independent complex antenna gains or receiver channel mismatch (and more generally the antenna coupling), e.g., [16, 17]. Much less studied are direction-dependent gains, which may include the individual antenna response, beamshape, and

angle-dependent propagation effects [18]. The latter case quickly leads to general models, e.g., to consider the array response matrix to have known gains and unknown phases [19, 20].

A related area of study (not immediately relevant for LOFAR) also considers estimating or updating the antenna locations along with the other parameters. Many calibration techniques essentially assume the presence of only a single calibration source, or even require a set of calibration sources which can be switched on or separately selected at will. This gives access to the individual uncalibrated array response vectors. Self calibration is an extension whereby DOA estimation (and hence source separation) is alternated with estimating the nuisance parameters. A problem rarely considered in array processing is to assume that many sources are simultaneously present, but that the source covariance matrix is known [21, 22]. This is a relevant assumption in radio astronomy, and LOFAR in particular.

2.4   Notation

Throughout this thesis the following notation will be used.

$\mathbf{I}, \mathbf{I}_L$: arbitrary size and $L \times L$ identity matrices respectively.

$\mathbf{0}, \mathbf{0}_{I,L}$: arbitrary and $I \times L$ matrix of zeros respectively.

$\mathbf{1}_L$: $L \times 1$ vector of ones.

$E\{\cdot\}$: expected value.

$\mathrm{Re}(\cdot), \mathrm{Im}(\cdot)$: real and imaginary parts, respectively.

$(\cdot)[i, j]$: (i, j)-th element of a matrix.

$\hat{(\cdot)}$: an estimated quantity.

$(\cdot)^{\mathrm{T}}, (\cdot)^{\mathrm{H}}$: transpose and Hermitian transpose respectively.

$(\cdot)^{\dagger}$: generalized inverse.

$\overline{(\cdot)}$: complex conjugate.

$\otimes, \circ$: Kronecker matrix product, Khatri-Rao (column-wise Kronecker) product.

$\mathrm{diag}(\cdot)$: extract diagonal, or build diagonal matrix.

$\mathrm{vec}(\cdot)$: Column scan a matrix to form a vector.

$|\cdot|, \angle$: element-wise absolute value and phase angle.

$\|\cdot\|_F$: Frobenius matrix norm.

3

Analysis of the LOFAR calibration problem [*]

This chapter presents a formal study of the parameter estimation problem for LOFAR calibration. A data model is proposed, and a Cramer-Rao lower bound (CRB) analysis is developed with a new general formulation to easily incorporate a variety of constraining signal models. It is shown that although the unconstrained direction dependent calibration problem is ambiguous, physically justifiable constraints can be applied in LOFAR to yield viable solutions. Use of a "compact core" of closely spaced array elements as part of the larger array is shown to significantly improve full array direction dependent calibration performance. Candidate algorithms are proposed and compared with the CRB.

This chapter is organized as follows. Section 3.1 introduces to problem of calibrating LOFAR. Section 3.2 gives the data model and problem statement. Section 3.3 derives the relevant CRBs. Section 3.4 proposes calibration algorithms, under various model assumptions and parameter constraints. Section 3.5 shows simulation results. Section 3.6 presents the conclusions of this chapter.

3.1   Introduction

In this chapter, we study the calibration of a large distributed sensor array. The problem can be phrased as direction-dependent calibration with fewer reference sources than array elements combined with hierarchical beamforming such that not all antenna crosscorrelations are available at a central location. The reference sources are signals of opportunity that are all simultaneously present. Without further assumptions, the array can not be

---

[*]The contents of this chapter have been published in *IEEE Transactions on Signal Processing* under the title "Self-Calibration for the LOFAR Radio Astronomical Array" by S. van der Tol, B.D. Jeffs, A.-J. van der Veen [23]. Prior conference publications providing partial results are [24, 25].

calibrated.

At the station level, each station is expected to form a well-defined beam into a desired direction–this requires accurate estimation of the complex gains of each antenna element in the station array. Available for this is the observed covariance matrix at each station, based on 1-second observations, and a table of the brightest sources in the sky, with known powers and locations. It is further assumed that the antenna elements have known locations and orientations, and "known" antenna patterns (as predicted by EM modeling). We estimate that each complex gain will have an accuracy of only 25 dB (relative to estimation error) [21], and this limits the knowledge and accuracy of the beamshape, in particular at the side lobes.

As with station calibration, bright point-like sky sources with known positions are used as "calibrator" references for ionospheric phase and gain estimation.

The initial station calibration should ensure that, within the mainlobe, the beamformer response of each station is sufficiently well known. Ionospheric variation across the field of view is gradual enough to permit a low order spatial smoothing model to fit observed perturbations to the known calibrator levels. The known beam response can aid direction dependent calibration and can be factored out so ionospheric gain and phase terms can be isolated.

However, in most observing scenarios there are multiple calibrators outside the mainlobe which are brighter than any source in the beam mainlobe, even after accounting for beamformer attenuation. Furthermore, sidelobe gain and phase responses vary rapidly with arrival angle and depend strongly on electronic instrument calibration variations. Sidelobe response to calibrators must thus be treated as an unknown random quantity which contributes to the direction dependent effect. The array must be accurately calibrated to these bright sources before their corrupting signals can be removed from the imaging array covariance data. This means that a LOFAR calibration algorithm must be capable of joint estimation of independent complex gain terms for every array element (station) and calibrator source combination. Moreover, at least for the first stages of calibration (used to remove bright sidelobe sources) this must be accomplished without the luxury of a known beam response. The algorithms and analysis presented here address this general case, where known beampatterns are not exploited.

At this stage in the LOFAR development there is significant uncertainty about how self calibration algorithms will perform. The radio astronomy community has a wealth of experience in successful synthesis array self calibration at higher frequencies [4] [5] [6] [13]. But neither the theoretical or practical bounds on calibration accuracy are well understood for arrays with thousands of antennas spread over a hundred kilometers in the presence of strong ionospheric perturbation. It is not clear whether extensions of existing algorithms will be adequate, and it is likely that new approaches and algorithms will be required [26]. For some observing conditions a sufficiently accurate calibration may be beyond fundamental limits of parameter estimation uncertainty. We propose to answer some of these questions with a thorough Cramer-Rao lower bound (CRB) analysis to determine limiting estimation error variance levels under various model assumptions.

In summary, compared to existing telescopes, LOFAR calibration has the following complications:

- The station beamshapes have significant side lobes; strong sources in the side lobes can dominate weak sources in the main lobe.

- Existing telescopes can calibrate assuming there is one or only a few bright sources in the field of view. For LOFAR, each omnidirectional antenna can see the full sky.

- Each station observes each source through a different patch of the ionosphere. It is easy to see from this that, without further assumptions, the array is not calibratable.

3.2   Data Model and Problem Definition

3.2.1   Signal Model

Each LOFAR station forms steered beams in 1 kHz wide subbands which track selected deep space objects while their apparent positions shift due to Earth rotation. A station beam is treated as a single directional element in the full LOFAR array for processing at the central location; there is no access to the individual elements. Assume all station beams in all subbands for the $J = 72$ stations are steered to the same point in the celestial sphere and that the observed signal is dominated by Q known, bright calibrator point sources. The $J \times 1$ observed array sample vector for the k-th subband centered at frequency f_k is

$$\mathbf{x}_k(n) \quad = \quad \sum_{q=1}^{Q} \mathbf{a}_{k,q}(n) s_{k,q}(n) + \boldsymbol{\eta}_k(n) \tag{3.1}$$

where $s_{k,q}(n)$ is the signal from the q-th calibrator source at time sample n and frequency f_k, $\mathbf{a}_{k,q}(n)$ is the array response vector for this source, and $\boldsymbol{\eta}_k(n)$ is the noise sample vector. $s_{k,q}(n)$ and $\boldsymbol{\eta}_k(n)$ are baseband complex envelope representations of zero mean wide sense stationary white Gaussian random processes sampled at the Nyquist rate. Elements of $\boldsymbol{\eta}_k(n)$ are statistically independent, as are signals from the Q sources.[1]

For simplicity of presentation all wave propagation is assumed to be non-polarized. In practice however antennas are grouped into orthogonal linear polarization pairs so full Stokes parameter outputs are available to enable observing polarization-specific scientific phenomena. Additionally, calibration parameter estimates must track the effect of ionospheric Faraday rotation. The non-polarized results presented in this paper are instructive, and extension to a more realistic model is straightforward using Jones matrix notation (cf. [27]), a dual-polarized vector in place of $s_{k,q}(n)$, and extending each array response vector $\mathbf{a}_{k,q}(n)$ to be a two column matrix.

Due to earth rotation the geometrical delay component of $\mathbf{a}_{k,q}(n)$ changes slowly with time, which is a critical feature exploited in synthesis imaging. Calibrator locations

---

[1]The subband processing in the actual instrument is slightly more subtle than presented here. The stations use 200 kHz subbands. At the central location these signals are time-shifted to compensate for the geometric delays in the look direction, and subsequently split into 1 kHz bins. As a result the narrowband model (3.1) holds for sources in the look direction, but may not quite hold for sources far outside the field of view. These sources will experience some phase smearing. This effect is not considered in the data model.

and intensities are known accurately from catalogues compiled in previous sky surveys. During calibration all other space signals are neglected due to their relative weakness, but of course their presence can bias the calibration solution.

Let N be the number of time samples in a short term integration (STI) interval. We assume that $\mathbf{a}_{k,q}(n)$ is (relatively) constant over such an interval, so that, for the m-th interval, $\mathbf{x}_k(n)$ is wide sense stationary over $(m-1)N \leq n \leq mN - 1$. A single STI autocovariance is defined as

$$\mathbf{R}_{k,m} = E\{\mathbf{x}_k(n)\,\mathbf{x}_k^{\mathrm{H}}(n)\} = \mathbf{A}_{k,m}\mathbf{\Sigma}_k\mathbf{A}_{k,m}^{\mathrm{H}} + \mathbf{\Lambda}_k, \tag{3.2}$$

where $\mathbf{R}_{k,m}$ has size $J \times J$,

$$\begin{aligned}
\mathbf{A}_{k,m} &= \left[\mathbf{a}_{k,1}((m-1)N), \cdots, \mathbf{a}_{k,Q}((m-1)N)\right] \\
\mathbf{\Sigma}_k &= \mathrm{diag}\{[\sigma_{k,1}^2, \cdots, \sigma_{k,Q}^2]\} \\
\mathbf{\Lambda}_k &= E\{\boldsymbol{\eta}_k(n)\,\boldsymbol{\eta}_k^{\mathrm{H}}(n)\} = \mathrm{diag}\{[\lambda_{k,1}^2, \cdots, \lambda_{k,J}^2]\}.
\end{aligned}$$

Here, $\sigma_{k,q}^2$ is the variance of the q-th calibrator source. Noise is assumed to be independent but not identically distributed across the array, and the noise variances $\lambda_{k,j}^2$ are unknown. In the radio astronomy literature, elements of $\mathbf{R}_{k,m}$ are called "visibilities" [4]. Each visibility represents the interferometric correlation along the baseline vector between the two corresponding array elements. The corresponding short term integration sample covariance estimate is

$$\hat{\mathbf{R}}_{k,m} = \frac{1}{N} \sum_{n=(m-1)N}^{mN-1} \mathbf{x}_k(n)\mathbf{x}_k^{\mathrm{H}}(n).$$

The array response matrix $\mathbf{A}_{k,m}$ can be factored into the product of a phase matrix $\mathbf{K}_{k,m}$ due entirely to the propagation delays associated with the array and source geometry, and a complex calibration gain matrix $\mathbf{G}_{k,m}$ which includes both source direction dependent ionospheric perturbations and electronic instrumentation gain errors,

$$\mathbf{A}_{k,m} = \mathbf{G}_{k,m} \odot \mathbf{K}_{k,m} \tag{3.3}$$

In the astronomical literature, the columns of $\mathbf{K}_{k,m}$, denoted by $\mathbf{k}_{k,m}^q$ $(q = 1, \cdots, Q)$, are often called the "Fourier kernel" and are given by

$$\begin{aligned}
\mathbf{k}_{k,m}^q &= \exp\{j\frac{2\pi f_k}{c}\mathbf{Z}^{\mathrm{T}}\mathbf{p}_{m,q}\} \\
\mathbf{Z} &= \left[[x_1, y_1, z_1]^{\mathrm{T}}, \cdots, [x_J, y_J, z_J]^{\mathrm{T}}\right].
\end{aligned}$$

where c is the speed of light, $[x_j, y_j, z_j]$ is the position vector for the j-th array element (station beam) and $\mathbf{p}_{m,q}$ is a unit length vector pointing in the direction of source q during STI snapshot m. Since $\mathbf{Z}$, $\mathbf{p}_{m,q}$, and the source power levels are all known to high accuracy for tabulated calibration sources, $\mathbf{K}_{k,m}$ and $\mathbf{\Sigma}_k$ are treated as known quantities.

3.2.2   Direction Dependent Calibration Formulation

The problem at hand is to estimate $\mathbf{G}_{k,m}$ given $\hat{\mathbf{R}}_{k,m}$ over a range of k and m. $\mathbf{G}_{k,m}$ is in general a $J \times Q$ full matrix of independent unknown complex gain parameters whose elements must be estimated to calibrate the array for imaging. A $(2JQ + J) \times 1$ real parameter vector containing all unknown terms is defined as

$$\boldsymbol{\theta}_{k,m} = [\text{vec}\{|\mathbf{G}_{k,m}|\}^{\text{T}}, \text{vec}\{\angle\mathbf{G}_{k,m}\}^{\text{T}}, \text{diag}\{\boldsymbol{\Lambda}_k\}]^{\text{T}}$$
$$= [(\boldsymbol{\gamma}_{k,m}^1)^{\text{T}}, \cdots (\boldsymbol{\gamma}_{k,m}^Q)^{\text{T}}, (\boldsymbol{\psi}_{k,m}^1)^{\text{T}}, \cdots (\boldsymbol{\psi}_{k,m}^Q)^{\text{T}}, \boldsymbol{\lambda}_k^{\text{T}}]^{\text{T}}, \qquad (3.4)$$

where $\boldsymbol{\gamma}_{k,m}^q$ is the q-th column of gain matrix $\boldsymbol{\Gamma} = |\mathbf{G}_{k,m}|$, $\boldsymbol{\psi}_{k,m}^q$ is the q-th column of phase matrix $\boldsymbol{\Psi} = \angle\mathbf{G}_{k,m}$, and $\boldsymbol{\lambda}_k = \text{diag}\{\boldsymbol{\Lambda}_k\}$. In contrast to LOFAR, the conventional synthesis imaging calibration problem at higher frequencies does not suffer from direction dependent ionospheric perturbations so in this case $\mathbf{G} = \mathbf{g}\mathbf{1}^{\text{T}}$ has the same gain vector $\mathbf{g}$ for each source, and (3.3) becomes $\mathbf{A} = \text{diag}\{\mathbf{g}\}\mathbf{K}$ [13]. In either case λ is a nuisance parameter which must be jointly estimated with $\mathbf{G}$.

Self calibration can be viewed as a covariance fitting problem. Substituting (3.3) into (3.2) and explicitly showing dependence on $\boldsymbol{\theta}$ yields the visibility measurement equation (ME) [28]

$$\text{ME}(\boldsymbol{\theta}_{k,m}) = (\mathbf{G}(\boldsymbol{\theta}_{k,m}) \odot \mathbf{K}_{k,m})\, \boldsymbol{\Sigma}_k \left(\mathbf{K}_{k,m}^{\text{H}} \odot \mathbf{G}^{\text{H}}(\boldsymbol{\theta}_{k,m})\right) + \boldsymbol{\Lambda}(\boldsymbol{\theta}_{k,m}).$$

For a single STI and subband [one (k, m)] the least squares calibration solution follows immediately as

$$\hat{\boldsymbol{\theta}}_{k,m} \quad = \quad \arg\min_{\boldsymbol{\theta}} \|\hat{\mathbf{R}}_{k,m} - \text{ME}(\boldsymbol{\theta})\|_F^2. \qquad (3.5)$$

Direct solution of (3.5) is not computationally practical. Furthermore it will be shown that without further constraints $\boldsymbol{\theta}_{k,m}$ is not identifiable through a single $\hat{\mathbf{R}}_{k,m}$. The estimation problem is ill posed and (3.5) yields ambiguous solutions due to source direction dependence. Fortunately the physics of LOFAR permit imposing structural constraints on individual $\mathbf{G}_{k,m}$ snapshots and/or across a range of time-frequency bins to regularize the problem, as will be shown in sections to follow.

It should also be mentioned that a maximum likelihood (ML) formulation of the LOFAR calibration problem is also easily expressed along the lines of the result in [13]. But, as is often the case, the ML approach does not yield a computationally tractable algorithm for this problem.

3.3   A Framework for Cramer-Rao Lower Bound Analysis

This section presents a general framework for CRB analysis of the source direction dependent calibration gain estimation problem. The approach allows for simple adaptation to a wide range of physically justifiable model assumptions, parameterizations, and signal constraints.

3.3.1   CRB for Unconstrained Calibration Parameters

Consider a set of array samples (J stations) observed over a time-frequency domain spanning $1 \leq k \leq K$ frequency bins and $1 \leq m \leq M$ non-overlapping N sample STI time windows. Stack these samples into an $KMJ \times N$ data matrix

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}_1(0) & \cdots & \mathbf{x}_1(N-1) \\ \vdots & & \vdots \\ \mathbf{x}_1((M-1)N) & \cdots & \mathbf{x}_1(MN-1) \\ \vdots & & \vdots \\ \mathbf{x}_k((m-1)N) & \cdots & \mathbf{x}_k(mN-1) \\ \vdots & & \vdots \\ \mathbf{x}_K((M-1)N) & \cdots & \mathbf{x}_K(MN-1) \end{bmatrix}.$$

Initially we consider the unstructured case where parameter vectors $\boldsymbol{\theta}_{k,m}$ from each time-frequency bin are distinct with no functional relationship and must all be estimated. These are stacked into a large parameter vector

$$\boldsymbol{\vartheta} = [\boldsymbol{\theta}_{1,1}^{\mathrm{T}}, \cdots, \boldsymbol{\theta}_{K,1}^{\mathrm{T}}, \cdots, \boldsymbol{\theta}_{1,M}^{\mathrm{T}}, \cdots \boldsymbol{\theta}_{K,M}^{\mathrm{T}}]^{\mathrm{T}}. \tag{3.6}$$

The corresponding stacked sample covariance is $\hat{\mathcal{R}} = \frac{1}{N}\mathbf{XX}^{\mathrm{H}}$. The underlying data model is given in (3.1). Due to Nyquist sampling of narrow passbands selected from the underlying continuous time broadband random signals, both $s_{k,q}(n)$ and $\boldsymbol{\eta}_k(n)$ are statistically independent with respect to bin indices k and m. All non zero correlations are spatial (i.e. with respect to station index j) and are due to phase delay across the narrowband array. Thus the true covariance $\mathcal{R} = E\{\hat{\mathcal{R}}\}$ has block diagonal form

$$\mathcal{R} = \begin{bmatrix} \mathbf{R}_{1,1} & & & & \\ & \ddots & & & \\ & & \mathbf{R}_{K,1} & & \\ & & & \ddots & \\ & & & & \mathbf{R}_{K,M} \end{bmatrix}$$

and $\mathcal{R}$ depends on $\boldsymbol{\vartheta}$ through (3.2), (3.3), and (3.4).

Consider an estimate of $\boldsymbol{\vartheta}$ based on an observation $\mathbf{X}$. The Cramer-Rao lower bound on the error variance for any unbiased estimator is given by the diagonal elements of

$$C = \frac{1}{N}\mathcal{M}^{-1}$$

evaluated at the true value of $\boldsymbol{\vartheta}$. Here, $\mathcal{M}$ is the Fisher information matrix, which for Gaussian data can be expressed as [29]

$$\mathcal{M} = \mathcal{J}^{\mathrm{H}}(\overline{\mathcal{R}}^{-1} \otimes \mathcal{R}^{-1})\mathcal{J} \tag{3.7}$$

where the Jacobian $\mathcal{J}$ is defined as

$$\mathcal{J} = \frac{\partial \mathrm{vec}(\mathcal{R})}{\partial \boldsymbol{\vartheta}^{\mathrm{T}}}.$$

The matrix $(\overline{\mathcal{R}}^{-1} \otimes \mathcal{R}^{-1})$ is usually very large. However the sparse block diagonal structure for $\mathcal{R}$ in turn makes $\mathcal{J}$ sparse and simplifies evaluation of (3.7). Thus $\mathcal{M}$ has the same block diagonal structure as $\mathcal{R}$, with

$$\mathbf{M}_{k,m} = \mathbf{J}_{k,m}^{\mathrm{H}} \left(\overline{\mathbf{R}}_{k,m}^{-1} \otimes \mathbf{R}_{k,m}^{-1} \right) \mathbf{J}_{k,m},$$

and

$$\mathbf{J}_{k,m} = \left. \frac{\partial \mathrm{vec}\left(\mathbf{R}_{k,m} \right)}{\partial \boldsymbol{\theta}_{k,m}^{\mathrm{T}}} \right|_{\boldsymbol{\theta} = \boldsymbol{\theta}_{k,m}^{\mathrm{true}}}.$$

With no assumed structure relating parameters across time or frequency, the subblocks of $\mathcal{M}$ are uncoupled and the CRB for some $\hat{\theta}_{k,m}$ can be computed from $\mathbf{M}_{k,m}$ independently from the other parameters. The entries $\mathbf{M}_{k,m}$ are evaluated as follows.

3.3.2   Closed Form for General Fisher Information

Using the parameter ordering from (3.4) each $\mathbf{M}_{k,m}$ can be partitioned into block form as

$$\mathbf{M}_{k,m} = \begin{bmatrix} \mathbf{M}_{\gamma_1\gamma_1} \cdots \mathbf{M}_{\gamma_1\gamma_Q} & \mathbf{M}_{\gamma_1\psi_1} \cdots \mathbf{M}_{\gamma_1\psi_Q} & \mathbf{M}_{\gamma_1\lambda} \\ \ddots & \ddots & \vdots \\ \mathbf{M}_{\gamma_Q\gamma_1} \cdots \mathbf{M}_{\gamma_Q\gamma_Q} & \mathbf{M}_{\gamma_Q\psi_1} \cdots \mathbf{M}_{\gamma_Q\psi_Q} & \mathbf{M}_{\gamma_Q\lambda} \\ \mathbf{M}_{\psi_1\gamma_1} \cdots \mathbf{M}_{\psi_1\gamma_Q} & \mathbf{M}_{\psi_1\psi_1} \cdots \mathbf{M}_{\psi_1\psi_Q} & \mathbf{M}_{\psi_1\lambda} \\ \ddots & \ddots & \vdots \\ \mathbf{M}_{\psi_Q\gamma_1} \cdots \mathbf{M}_{\psi_Q\gamma_Q} & \mathbf{M}_{\psi_Q\psi_1} \cdots \mathbf{M}_{\psi_Q\psi_Q} & \mathbf{M}_{\psi_Q\lambda} \\ \mathbf{M}_{\lambda\gamma_1} \cdots \mathbf{M}_{\lambda\gamma_Q} & \mathbf{M}_{\lambda\psi_1} \cdots \mathbf{M}_{\lambda\psi_Q} & \mathbf{M}_{\lambda\lambda} \end{bmatrix}. \qquad (3.8)$$

The closed form representation for these submatrices is shown in the Appendix to be

$$\mathbf{M}_{\gamma_p\gamma_q} = 2\sigma_p^2\sigma_q^2 \mathrm{Re}\left\{ (\boldsymbol{\Phi}_p^{\mathrm{T}}\overline{\mathbf{R}}^{-1}\overline{\boldsymbol{\Phi}}_q)(\mathbf{a}_p^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{a}_q) + (\boldsymbol{\Phi}_p^{\mathrm{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{a}}_q)(\mathbf{a}_p^{\mathrm{H}}\mathbf{R}^{-1}\boldsymbol{\Phi}_q) \right\} \qquad (3.9)$$

$$\mathbf{M}_{\psi_p\psi_q} = 2\sigma_p^2\sigma_q^2 \mathrm{Re}\left\{ (\mathbf{E}_p\overline{\mathbf{R}}^{-1}\overline{\mathbf{E}}_q)(\mathbf{a}_p^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{a}_q) - (\mathbf{E}_p\overline{\mathbf{R}}^{-1}\overline{\mathbf{a}}_q)(\mathbf{a}_p^{\mathrm{H}}\overline{\mathbf{R}}^{-1}\mathbf{E}_q) \right\} \qquad (3.10)$$

$$\mathbf{M}_{\lambda\lambda} = \overline{\mathbf{R}}^{-1} \odot \mathbf{R}^{-1} \qquad (3.11)$$

$$\mathbf{M}_{\gamma_p\psi_q} = 2\sigma_p^2\sigma_q^2 \mathrm{Im}\left\{ (\boldsymbol{\Phi}_p^{\mathrm{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{E}}_q)(\mathbf{a}_p^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{a}_q) + (\boldsymbol{\Phi}_p^{\mathrm{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{a}}_q)(\mathbf{a}_p^{\mathrm{H}}\mathbf{R}^{-1}\mathbf{E}_q) \right\} \qquad (3.12)$$

$$\mathbf{M}_{\gamma_q\lambda} = 2\sigma_q^2 \mathrm{Re}\left\{ \boldsymbol{\Phi}_q\overline{\mathbf{R}}^{-1} \circ \mathbf{a}_q^{\mathrm{H}}\mathbf{R}^{-1} \right\} \qquad (3.13)$$

$$\mathbf{M}_{\psi_q\lambda} = -2\sigma_q^2 \mathrm{Im}\left\{ \mathbf{E}_q\overline{\mathbf{R}}^{-1} \circ \mathbf{a}_q^{\mathrm{H}}\mathbf{R}^{-1} \right\} \qquad (3.14)$$

where all terms have implied k, m subscripts and

$$
\begin{aligned}
\mathbf{E}_q &= \operatorname{diag}\left(\boldsymbol{\gamma}_q \odot \mathbf{k}_q\right) \\
\boldsymbol{\Phi}_q &= \operatorname{diag}\left(\boldsymbol{\phi}_q \odot \mathbf{k}_q\right) \\
\boldsymbol{\phi}_q &= \exp(j\boldsymbol{\psi}_q).
\end{aligned}
$$

Since $\mathbf{M}_{\boldsymbol{\gamma}_p \lambda}$ and $\mathbf{M}_{\boldsymbol{\psi}_p \lambda}$ are nonzero, λ is coupled with the parameters of interest and should be jointly estimated.

   As will be shown later, in the general case considered here $\mathbf{M}_{k,m}$ is singular. Without introducing constraints on the parameters, the array is not calibratable.

3.3.3   CRB for Constrained Parameters

The following sections discuss some scenarios where the degrees of freedom in $\boldsymbol{\vartheta}$ can be reduced by physically justifiable constraining models. These impose structure in $\mathcal{R}$ which is key to solving the calibration problem. At this stage in LOFAR development there is much activity in identifying appropriate models of ionospheric perturbation effects which can be incorporated into self calibration algorithms (see chapter 4). For example, when the ionosphere is relatively time stable, $\boldsymbol{\theta}_{k,m}$ may vary smoothly over k and m according to some low order interpolation function. In such cases a lower dimensional parameter vector $\boldsymbol{\rho}$ can represent all the required degrees of freedom over the entire domain of k and m.

   Let $\boldsymbol{\vartheta} = f(\boldsymbol{\rho})$ where $f(\cdot)$ is some functional relationship describing a constraint on $\boldsymbol{\vartheta}$ corresponding to an appropriate physical model. It is assumed that $\boldsymbol{\vartheta}$ is an overdetermined parameterization and that the underlying distribution for $\mathbf{x}_k(n)$ is fully determined by $\boldsymbol{\rho}$. Under these conditions the Fisher information computed for $\boldsymbol{\rho}$ yields the CRB.

   Define the constrained Fisher information matrix as

$$
\mathcal{M}_{\boldsymbol{\rho}} = \mathcal{J}_{\boldsymbol{\rho}}^{\mathrm{H}}\left(\overline{\mathcal{R}}^{-1} \otimes \mathcal{R}^{-1}\right)\mathcal{J}_{\boldsymbol{\rho}} \tag{3.15}
$$

where

$$
\mathcal{J}_{\boldsymbol{\rho}} = \frac{\partial \operatorname{vec}(\mathcal{R})}{\partial \boldsymbol{\rho}^{\mathrm{T}}} = \frac{\partial \operatorname{vec}(\mathcal{R})}{\partial \boldsymbol{\vartheta}^{\mathrm{T}}}\frac{\partial f(\boldsymbol{\rho})}{\partial \boldsymbol{\rho}^{\mathrm{T}}} =: \mathcal{J}\mathcal{F} \tag{3.16}
$$

which follows by the chain rule since $\boldsymbol{\vartheta} = f(\boldsymbol{\rho})$. $\mathcal{F}$ can be partitioned as

$$
\mathcal{F} = \left[\mathbf{F}_{1,1}^{\mathrm{T}}, \cdots, \mathbf{F}_{K,1}^{\mathrm{T}}, \cdots, \mathbf{F}_{1,M}^{\mathrm{T}}, \cdots, \mathbf{F}_{K,M}^{\mathrm{T}}\right]^{\mathrm{T}}
$$

where

$$
\mathbf{F}_{k,m} = \frac{\partial \boldsymbol{\theta}_{k,m}}{\partial \boldsymbol{\rho}^{\mathrm{T}}}. \tag{3.17}
$$

$\boldsymbol{\rho}$ is common to all time-frequency bins and is not indexed by k or m. Specifying the $\mathbf{F}_{k,m}$ establishes the required structural constraints. Equation (3.15) can be written as

$$
\mathcal{M}_{\boldsymbol{\rho}} = \mathcal{F}^{\mathrm{H}}\left[\mathcal{J}^{\mathrm{H}}\left(\overline{\mathcal{R}}^{-1} \otimes \mathcal{R}^{-1}\right)\mathcal{J}\right]\mathcal{F} = \mathcal{F}^{\mathrm{H}}\mathcal{M}\mathcal{F},
$$

and since $\mathcal{M}$ is block diagonal

$$\mathcal{M}_{\boldsymbol{\rho}} = \sum_{k=1}^{K} \sum_{m=1}^{M} \mathbf{F}_{k,m}^{\text{H}} \mathbf{M}_{k,m} \mathbf{F}_{k,m}. \tag{3.18}$$

With (3.18) one may compute a CRB for the constrained parameter vector $\boldsymbol{\rho}$ using the unconstrained general form Fisher matrices given by (3.8) and constraint Jacobians from (3.17).

As an illustrative example consider the simplest time-frequency smoothing function where calibration parameters are constant over k, m. In this case we can choose $\boldsymbol{\rho} = \boldsymbol{\theta}_{1,1} = \boldsymbol{\theta}_{k,m} \ \forall (k, m)$, and $\mathbf{F}_{k,m} = \mathbf{I}$. To avoid a singular $\mathcal{M}_{\boldsymbol{\rho}}$ an intrinsic bulk phase ambiguity must be resolved with an additional constraint. Since due to its Hermitian product form $\mathbf{R}_{k,m}$ is unaffected by multiplying any column of $\mathbf{G}_{k,m}$ by a unit modulus scalar, $\angle \mathbf{G}_{k,m}$ can only be known to within one arbitrary phase factor per column. The excess degrees of freedom in $\boldsymbol{\theta}_{k,m}$ can be removed by eliminating the first element of each phase vector $\boldsymbol{\psi}_{k,m}^{q}$ in $\boldsymbol{\theta}_{k,m}$. This constraint is imposed by setting

$$\boldsymbol{\rho} = \mathbf{S}_{L}^{\mathbf{b}} \boldsymbol{\theta}_{1,1} = \mathbf{S}_{L}^{\mathbf{b}} \boldsymbol{\theta}_{k,m}, \quad \forall k, m,$$

where $L = (2Q + 1)J$ is the number of coefficients in $\boldsymbol{\theta}_{k,m}$, and the selection matrix $\mathbf{S}_{L}^{\mathbf{b}}$ is formed by deleting the columns from $\mathbf{I}_{L}$ with indices $\mathbf{b} = [b_1, \cdots, b_Q]^{\text{T}}$, $b_q = (q - 1 + Q)J + 1$. This eliminates from $\boldsymbol{\rho}$ the station $j = 1$ phase parameter for each of the Q calibrator sources, forcing the first row of $\mathbf{G}$ to be real. It follows that

$$\mathbf{F}_{k,m} = \mathbf{S}_{L}^{\mathbf{b}}. \tag{3.19}$$

The resulting bounds are evaluated in Section 3.5, along with the performance of the estimation algorithms presented next.

3.4 Calibration Methods

3.4.1 Single Snapshot Calibration

Above approximately 400 MHz it is possible for conventional astronomical synthesis imaging arrays to estimate calibration gains from a single "snapshot" STI sample covariance realization $\hat{\mathbf{R}}_{k,m}$ [13] [4]. This is useful for start-up of a tracking calibration algorithm or to make quick look snapshot images. In general this is not possible for LOFAR due to directionally dependent ionospheric perturbations.

To illustrate this fact note that for any $Q \times Q$ unitary matrix $\mathbf{U}$, equation (3.2) can be rewritten as

$$\begin{aligned} \mathbf{R} &= \mathbf{A} \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{U} \mathbf{U}^{\text{H}} \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{A}^{\text{H}} + \boldsymbol{\Lambda} \\ &= \mathbf{A} \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{U} \boldsymbol{\Sigma}^{-\frac{1}{2}} \boldsymbol{\Sigma} (\boldsymbol{\Sigma}^{-\frac{1}{2}})^{\text{H}} \mathbf{U}^{\text{H}} \boldsymbol{\Sigma}^{\frac{1}{2}} \mathbf{A}^{\text{H}} + \boldsymbol{\Lambda} \\ &= \tilde{\mathbf{A}} \boldsymbol{\Sigma} \tilde{\mathbf{A}}^{\text{H}} + \boldsymbol{\Lambda}, \end{aligned} \tag{3.20}$$

Full LOFAR array geometry for 72 stations   Closeup of central core of 32 stations
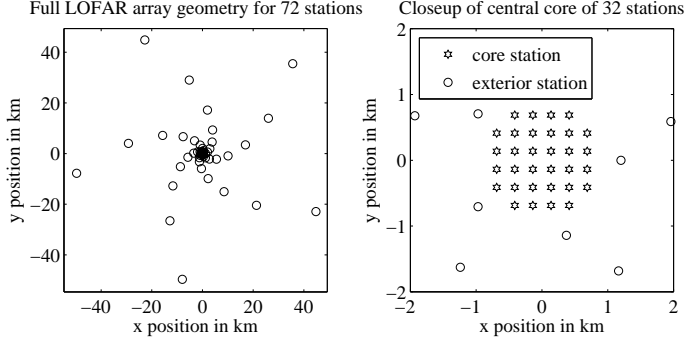
Figure 3.1: *Possible geometry for the full LOFAR array (left) and the compact central core (right). Each circle or star represents a LOFAR station which acts as a single beamformed directional sensor element in the full array. Plans in 2004 were for five exponentially spaced spiral arms of eight exterior stations each and a compact core of 32 stations.*

where $\tilde{\mathbf{A}} = \mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma}^{-\frac{1}{2}}$. Since $\mathbf{U}\mathbf{U}^{\mathrm{H}} = \mathbf{I}$, it is not visible in $\mathbf{R}$. However (3.20) has the same structure as (3.2) but with a different effective array response, $\tilde{\mathbf{A}}$. Thus each choice of $\mathbf{U}$ leads to a different calibration solution $\tilde{\mathbf{G}}$, namely (using $\tilde{\mathbf{A}} = \tilde{\mathbf{G}} \odot \mathbf{K}$),

$$\tilde{\mathbf{G}} = (\mathbf{A}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}\boldsymbol{\Sigma}^{-\frac{1}{2}}) \odot \mathbf{K}^{\odot-1} \tag{3.21}$$

where $(\cdot)^{\odot-1}$ denotes element-wise inverse. This shows that, without adding constraints, $\mathbf{A}$ is not identifiable from a single $\hat{\mathbf{R}}_{k,m}$. (This problem is not present in the classical direction-independent calibration problem, where $\mathbf{G} = \mathbf{g}\mathbf{1}^{\mathrm{T}}$, or $\mathbf{A} = \mathrm{diag}\{\mathbf{g}\}\mathbf{K}$ [13].)

In the next subsection, we introduce a physically justifiable constraint based on the unique LOFAR array geometry which resolves this ambiguity so that with sufficient SNR a calibration can be computed already from a single snapshot $\hat{\mathbf{R}}_{k,m}$. Subsequently, in Section 3.4.3, we consider multiple snapshots and make assumptions on the time evolution of the ionosphere.

3.4.2   Exploiting the Compact Core LOFAR Geometry

The unknown calibration gains/phases can be attributed to perturbations due to (*a*) the propagation through the ionosphere, and (*b*) the receiver electronics. The ionosphere mostly introduces propagation delays, i.e. it can be modeled as a random phase sheet, with gains that are approximately direction independent over a station main beam. Similarly, the electronic gains and phases are independent of the directions to the calibration sources, but do differ from station to station.

The planned geometry for LOFAR as shown in Fig. 3.1 includes a central core of J_c closely packed stations. As shown in Fig. 3.2, the core subarray is operating in a regime 3 (see section 2.2.2) where the station beam fields of view overlap on the ionosphere (approximately 300 km above the array). These beam mainlobe "footprints" are much
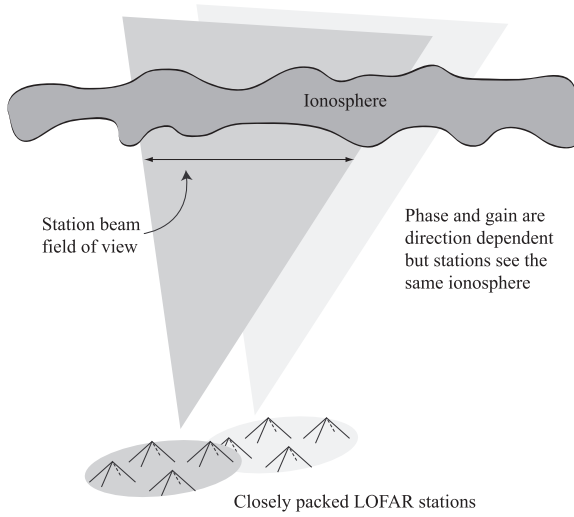
Figure 3.2: *Calibration scenario for closely spaced LOFAR central core stations. Due to beam overlap at ionospheric altitude, each station sees the same direction dependence. (After C. Lonsdale)*

larger than, and the total subarray aperture is much smaller than the ionospheric irregularity scale [9]. Thus each core station observes a given celestial source through the same patch of ionosphere and sees a common gain-phase perturbation, i.e., the core subarray sees a coherent scene without direction dependence.

Because they are common, the ionospheric phases cancel out when computing the correlations $\mathbf{R}_{k,m}$ for the core subarray. What remains is the direction-independent ionospheric gains, and the gains/phases of the receiver electronics. The corresponding core gain matrix $\mathbf{G}_c$ can be modeled as $\mathbf{G}_c = \mathbf{g}_c \mathbf{1}^{\mathrm{T}}$. This leads to a calibration problem comparable to the typical synthesis imaging situation at higher frequencies (or similar to station calibration).

For the rest of the array, i.e., for the $J_e = J - J_c$ stations exterior to the core, both the field of view and all inter-element baselines are greater than the ionospheric irregularity scale, as shown in Fig. 2.1(d). For these stations the corresponding gain matrix $\mathbf{G}_e$ is best modeled as a full matrix.

Under these assumptions, equation (3.3) becomes

$$\mathbf{A} = \mathbf{G} \odot \mathbf{K} \;\; = \;\; \left[\begin{array}{c} \mathbf{G}_c \odot \mathbf{K}_c \\ \mathbf{G}_e \odot \mathbf{K}_e \end{array} \right] = \left[\begin{array}{c} \mathrm{diag}\{\mathbf{g}_c\}\mathbf{K}_c \\ \left[\mathbf{g}_e^1 \odot \mathbf{k}_e^1, \cdots, \mathbf{g}_e^Q \odot \mathbf{k}_e^Q \right] \end{array} \right],$$

(the implied (k,m) indices are omitted). To bring all calibration gain vectors to be estimated into a single indexing scheme, let $q = 0$ refer to the central core, and define

$$\tilde{\boldsymbol{g}}_q = \left\{ \begin{array}{ll} \mathbf{g}_c, & \text{for } q = 0, \\ \mathbf{g}_e^q, & \text{for } 1 \leq q \leq Q \end{array} \right.$$

$$\boldsymbol{\rho} = [|\tilde{\boldsymbol{g}}_0^{\mathsf{T}}|, \cdots, |\tilde{\boldsymbol{g}}_Q^{\mathsf{T}}|, \angle\tilde{\boldsymbol{g}}_0^{\mathsf{T}}, \cdots, \angle\tilde{\boldsymbol{g}}_Q^{\mathsf{T}}, \boldsymbol{\lambda}^{\mathsf{T}}]^{\mathsf{T}}. \tag{3.22}$$

Here, $\tilde{\boldsymbol{g}}_0$ corresponding to the single core subarray gain vector is length J_c, while the $\tilde{\boldsymbol{g}}_q$ are length J_e.

The following constraint Jacobian expresses the relationship between $\boldsymbol{\rho}$ in (3.22) and $\boldsymbol{\theta}_{k,m}$ in (3.4). It also constrains the bulk phase ambiguity as does (3.19), by omitting the phase on the first external station.

$$\mathbf{F}_{k,m} = \begin{bmatrix} \mathbf{1}_Q \otimes \begin{bmatrix} \mathbf{I}_{J_c} \\ \mathbf{0}_{J_e, J_c} \end{bmatrix} & \mathbf{I}_Q \otimes \begin{bmatrix} \mathbf{0}_{J_c, J_e} \\ \mathbf{I}_{J_e} \end{bmatrix} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1}_Q \otimes \begin{bmatrix} \mathbf{I}_{J_c} \\ \mathbf{0}_{J_e, J_c} \end{bmatrix} & \mathbf{I}_Q \otimes \begin{bmatrix} \mathbf{0}_{J_c, J_e-1} \\ \mathbf{S}_{J_e}^{\mathbf{b}} \end{bmatrix} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_M \end{bmatrix} \tag{3.23}$$

where $\mathbf{b} = [1]$. Using (3.23) in (3.18) for a single snapshot (k, m) gives $\mathcal{M}_{\boldsymbol{\rho}} = \mathbf{F}_{k,m}^{\mathrm{H}} \mathbf{M}_{k,m} \mathbf{F}_{k,m}$, and yields a closed form CRB for a LOFAR array with a central core. This has been used to evaluate the CRB for a wide range of scenarios, and leads to the following observations:

1. $\mathbf{M}_{\boldsymbol{\rho}}$ is typically singular when $J_c < Q - 1$.

2. When $J_c \geq Q - 1$ the full array, including the fully direction dependent ionospheric gains in the exterior stations can be reliably calibrated with a single snapshot sample covariance.

Currently, the number of central compact core stations is planned as $J_c = 32$. Thus, single snapshot calibration exploiting the core configuration is suitable if Q represents a small number of bright calibrator sources, i.e., for initial coarse calibration.

3.4.3 Exploiting Frequency-Time Diversity

We will now consider the use of multiple snapshots. The ionospheric parameters are approximately constant over a block of (k, m) values covering 10 seconds and 500 kHz. Due to Earth rotation and frequency dependence, $\mathbf{K}_{k,m}$ varies sufficiently over this block so that, even if the individual $\mathbf{M}_{k,m}$ are singular, the sum in (3.18) produces a full rank $\mathcal{M}_{\boldsymbol{\rho}}$ and a relatively low CRB.

For regions larger than a 10 s $\times$ 500 kHz block, $\mathbf{G}_{k,m}$ generally varies smoothly. It is therefore unnecessary to compute independent estimates for each $\boldsymbol{\theta}_{k,m}$. A low order smoothing function can describe the significant variations with fewer parameters and thus lower estimation error variance. For example, we can use a matrix polynomial in f, t and coefficient vector $\boldsymbol{\rho}$ for the phase matrix $\boldsymbol{\Psi}$ as

$$\boldsymbol{\Psi}(\boldsymbol{\rho}, f, t) = \mathbf{T}_1 + \mathbf{T}_2 f + \mathbf{T}_3 t + \mathbf{T}_4 f t + \mathbf{T}_5 f^2$$

which consists of $D = 5$ terms, using powers for f as $\{\kappa_1, \cdots, \kappa_5\} = \{0, 1, 0, 1, 2\}$ and for t as $\{\mu_1, \cdots, \mu_5\} = \{0, 0, 1, 1, 0\}$, and with the $\mathbf{T}_d$ phase coefficient matrices of size $J \times Q$.

More generally, we can model the magnitude and phase matrix at STI frequency-time bin (k, m) as

$$\mathbf{G}_{k,m}(\boldsymbol{\rho}) = \left[\boldsymbol{\Gamma}(\boldsymbol{\rho}, f, t) \odot \exp\{j\boldsymbol{\Psi}(\boldsymbol{\rho}, f, t)\}\right]\Big|_{f=f_k, t=t_k} \tag{3.24}$$

$$\boldsymbol{\Gamma}(\boldsymbol{\rho}, f, t) = \sum_{d=1}^{D} \mathbf{Y}_d f^{\kappa_d} t^{\mu_d}, \quad \boldsymbol{\Psi}(\boldsymbol{\rho}, f, t) = \sum_{d=1}^{D} \mathbf{T}_d f^{\kappa_d} t^{\mu_d}$$

$$\boldsymbol{\rho} = \text{vec}\,[\mathbf{Y}_1, \cdots \mathbf{Y}_D, \mathbf{T}_1, \cdots \mathbf{T}_D, \boldsymbol{\lambda}]. \tag{3.25}$$

$\mathbf{Y}_d$ and $\mathbf{T}_d$ are the $J \times Q$ gain and phase coefficient matrices, to be estimated by calibration.

The dual (gain and phase) polynomial model combined with the bulk phase ambiguity resolution used in (3.19) results in the following constraint Jacobian:

$$\mathbf{F}_{k,m} = \begin{bmatrix} f_k^{\kappa_1} t_m^{\mu_1} \mathbf{I}_L & \cdots & f_k^{\kappa_D} t_m^{\mu_D} \mathbf{I}_L & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & f_k^{\kappa_1} t_m^{\mu_1} \mathbf{B} & \cdots & f_k^{\kappa_D} t_m^{\mu_D} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{I}_J \end{bmatrix} \tag{3.26}$$

where $L = JQ$ and $\mathbf{B} = \mathbf{I}_Q \otimes \mathbf{S}_J^{\mathbf{b}}$, $\mathbf{b} = [1]$. It is straightforward to combine this with the assumption of a central core geometry (Section 3.4.2). In this case, $\mathbf{F}_{k,m}$ is given by the product of equations (3.23) and (3.26), but using $L = QJ_e + J_c$ and $\mathbf{B} = \mathbf{I}_{Q(J_e-1)+J_c}$.

A least squares solution for $\boldsymbol{\rho}$ can be expressed as

$$\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} \|\hat{\mathbf{R}}_{k,m} - \text{ME}_{k,m}(\boldsymbol{\rho})\|_F^2 \tag{3.27}$$

where

$$\text{ME}_{k,m}(\boldsymbol{\rho}) = (\mathbf{G}_{k,m}(\boldsymbol{\rho}) \odot \mathbf{K}_{k,m}) \boldsymbol{\Sigma}_k \left(\mathbf{K}_{k,m}^{\text{H}} \odot \mathbf{G}_{k,m}(\boldsymbol{\rho})^{\text{H}}\right) + \boldsymbol{\Lambda}_k(\boldsymbol{\rho}). \tag{3.28}$$

Although the polynomial model dramatically reduces parameter degrees of freedom and (3.26) yields a low CRB, a direct implementation of (3.27) is computationally impractical. An iterative search algorithm is required but its convergence performance is poor. The continuous phase polynomial $\boldsymbol{\Psi}(\boldsymbol{\rho}, f, t)$ is ambiguous to integer multiples of 2π at every evaluation point, i.e. for every combination of station j, source q, f_k, and t_m. This introduces many local minima so that a good initial estimate for $\boldsymbol{\rho}$ is required. The following two sections present algorithms which address these problems.

3.4.4   The Peeling Algorithm

The current leading candidate algorithm for LOFAR calibration was introduced in [26] and has been dubbed "Peeling" due its sequential approach of successively calibrating on one bright source at a time followed by removing (peeling) that source's contribution from the observed sample covariances, $\hat{\mathbf{R}}_{k,m}$. Peeling is based on three basic simplifying assumptions:

- Joint estimation for parameters of all Q calibrators sources can be approximated with a series of single source calibration problems, in descending order of source brightness.

- Calibration gains vary slowly and smoothly over time and frequency. Consequently, over some span of K_b frequency bins and M_b time bins called a "block," $\mathbf{G}_{k,m}$ is approximately constant. This block indexed by $(\tilde{k}, \tilde{m})$ includes all frequency-time (STI) bins in the set

$$\mathcal{B}_{\tilde{k},\tilde{m}} = \{(k,m) : \tilde{k}K_b \le k \le (\tilde{k}+1)K_b - 1, \ \tilde{m}M_b \le m \le (\tilde{m}+1)M_b - 1\}$$

  The evolution of the ionospheric gains over several blocks (a "domain") is described by a polynomial model as in (3.25).

- Within a block, the variations in $\mathbf{K}_{k,m}$ (also known as fringe rotations) due to Earth rotation and frequency change are large. Source powers $\sigma_{k,q}^2$ are constant over k within a block.

Peeling in [26] does not use the central core geometry assumption.

Let $\boldsymbol{\rho}$ be a minimal parameterization for the calibration parameters. Define $\boldsymbol{\rho}_q$ as the subvector of $\boldsymbol{\rho}$ corresponding to the parameters for source q, i.e., corresponding to the q-th columns from $\mathbf{Y}_1, \cdots, \mathbf{Y}_D$ and $\mathbf{T}_1, \cdots, \mathbf{T}_D$ in the polynomial model (3.25). The corresponding gain vector for STI bin (k,m) is given by the vector polynomial $\mathbf{g}_{k,m}^q = \mathbf{g}_{k,m}(\boldsymbol{\rho}_q)$, which is also obtained by by retaining only the q-th columns from $\boldsymbol{\Gamma}(\boldsymbol{\rho}, f, t)$ and $\boldsymbol{\Psi}(\boldsymbol{\rho}, f, t)$ in (3.24). No superscript q is used in $\mathbf{g}_{k,m}(\boldsymbol{\rho}_q)$ since all column-wise polynomials are identical except for the source dependent coefficients in each $\boldsymbol{\rho}_q$. Similarly, $\mathbf{k}_{k,m}^q$ denotes the q-th column of $\mathbf{K}_{k,m}$, and contains the geometric phase delays of source q. Finally, $\hat{\boldsymbol{\rho}}_p$ will denote the current parameter vector estimate for a single source p.

Assuming the Q sources are ordered in descending brightness, an I pass peeling algorithm based on [26] is given by

1. *Initialize:* source index $q = 1$, pass index $i = 1$, and parameter vector $\hat{\boldsymbol{\rho}}_p = \mathbf{0}$ for $1 \le p \le Q$.

2. *Update the residuals (peel):* Over all (k,m) covering all blocks in the domain, subtract from each sample covariance the current best estimates (based on $\hat{\boldsymbol{\rho}}_p$, $1 \le p \le Q$, $p \ne q$) of contributions from all except the q-th source:

$$\hat{\mathbf{V}}_{k,m,q} = \hat{\mathbf{R}}_{k,m} - \sum_{\substack{p=1 \\ p \ne q}}^{Q} (\mathbf{g}_{k,m}(\hat{\boldsymbol{\rho}}_p) \odot \mathbf{k}_{k,m}^p) \sigma_{k,p}^2 (\mathbf{g}_{k,m}(\hat{\boldsymbol{\rho}}_p) \odot \mathbf{k}_{k,m}^p)^{\mathrm{H}}.$$

$\hat{\mathbf{V}}_{k,m,q}$ is an estimate of the visibility (covariance) matrix contribution to $\hat{\mathbf{R}}_{k,m}$ from source q. The term under the summation is a single source version of (3.28). The noise covariance $\boldsymbol{\lambda}$ is neglected.

3. *Phase center and average:* For each $\hat{\mathbf{V}}_{k,m,q}$, cancel the phase rotation due to the geometric delay term $\mathbf{k}_{k,m}^q$ in the visibility contribution from source q. Then average over (k,m) inside a block $\mathcal{B}_{\tilde{k},\tilde{m}}$ to attenuate the other, non-centered, sources,

$$\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q} = \frac{1}{K_b M_b} \sum_{(k,m)\in\mathcal{B}_{\tilde{k},\tilde{m}}} \text{diag}\{\overline{\mathbf{k}}_{k,m}^q\} \hat{\mathbf{V}}_{k,m,q} \, \text{diag}\{\mathbf{k}_{k,m}^q\}. \qquad (3.29)$$

4. *Estimate polynomial coefficients:* With some abuse of notation, let the subscript $(\tilde{k},\tilde{m})$ on $\mathbf{g}^q_{\tilde{k},\tilde{m}}$ denote selecting the (k,m) frequency-time bin in the center of block $(\tilde{k},\tilde{m})$, and likewise for $\sigma^2_{\tilde{k},q}$. Then assume a single-source model and estimate the polynomial coefficients for source q as

$$\hat{\boldsymbol{\rho}}_q \;=\; \arg\min_{\boldsymbol{\rho}_q} \sum_{(\tilde{k},\tilde{m})} \left\| \mathbf{L} \odot \left(\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q} - \sigma^2_{\tilde{k},q} \mathbf{g}_{\tilde{k},\tilde{m}}(\boldsymbol{\rho}_q) \left(\mathbf{g}_{\tilde{k},\tilde{m}}(\boldsymbol{\rho}_q) \right)^{\text{H}} \right) \right\|_{\text{F}}^2 .$$

$\mathbf{L}$ is a masking matrix of ones below the diagonal and zeros elsewhere which is used to avoid fitting to diagonal terms from $\boldsymbol{\Lambda}$. This problem is solved using a general least squares solver.

5. *Iterate* for $q = 1, \cdots, Q$ and do this for I passes.

We have found that using multiple passes (e.g. $2 \leq I \leq 5$) reduces bias in $\hat{\boldsymbol{\rho}}_q$ which arises when the averaging over a block in step 3 produces insufficient attenuation of the non centered sources. Contamination in the single source fit in step 4 occurs because $\mathcal{V}_{\tilde{k},\tilde{m},q}$ has contributions from more than the centered source. The next section presents a more direct method of reducing this bias.

3.4.5   Demixing Calibrator Cross Contamination

The purpose of steps 2 and 3 is to form a single-source approximation of the problem. Ideally $\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q}$ is equal to the true single source phase centered visibility $\mathcal{V}_{k,m,q}$ for sample (k,m) at the center of block $(\tilde{k},\tilde{m})$. Assuming the gains $\mathbf{g}_{k,m}(\boldsymbol{\rho}_q)$ and source powers $\sigma^2_{k,q}$ are constant within the block gives

$$\mathcal{V}_{\tilde{k},\tilde{m},q} = \sigma^2_{k,q} \mathbf{g}^q_{k,m}(\boldsymbol{\rho}_q) \left(\mathbf{g}_{k,m}(\boldsymbol{\rho}_q) \right)^{\text{H}}, \qquad \forall (k,m) \in \mathcal{B}_{\tilde{k},\tilde{m}}. \qquad (3.30)$$

(the additive noise is ignored). Initially, when estimating the parameters for source q there are no available estimates for sources $q+1$ to Q, so their contribution cannot be subtracted in step 2. Averaging in step 3 is then not sufficient to reduce bias down to the noise level and multiple iterations are necessary.

In this section we develop an unbiased estimator, $\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q}$, so that even on the first pass $E\{\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q}\} \approx \mathcal{V}_{\tilde{k},\tilde{m},q}$. The algorithm works on a per block basis, so for notational simplicity the block indices $(\tilde{k},\tilde{m})$ will be dropped, and we take the ranges $k = 1, \cdots, K_b$, $m = 1, \cdots, M_b$ (a single block).

Consider estimating $\boldsymbol{\rho}_q$ during the first peeling pass. Estimates $\hat{\boldsymbol{\rho}}_p$ for $1 \leq p < q$ will have been previously computed, and we assume that the corresponding sources are peeled without bias in step 2. We now seek an estimate $\hat{\boldsymbol{\rho}}_q$ which is unbiased by the presence of sources $q + 1$ to Q in $\hat{\mathbf{R}}_{k,m}$. These sources have not yet been peeled since at this stage $\hat{\boldsymbol{\rho}}_p = \mathbf{0}$ for $p = q + 1, \cdots, Q$. The expected value of entry i, j from $\hat{\mathbf{V}}_{k,m,q}$ in step 2 can then be expressed as (for $i, j = 1, \cdots, J$)

$$E\left\{\hat{\mathbf{V}}_{k,m,q}[i, j]\right\} = \sum_{p=q}^{Q} \mathbf{k}_{k,m}^{p}[i]\, \overline{\mathbf{k}}_{k,m}^{p}[j]\, \nu_{p,ij} \tag{3.31}$$

$$\nu_{p,ij} = \sigma_p^2 \mathbf{g}_p[i]\, \overline{\mathbf{g}}_p[j]$$

where $\mathbf{g}_q = \mathbf{g}_{k,m}(\boldsymbol{\rho}_q)$ and $\sigma_q^2 = \sigma_{k,q}^2$ are constant within a block. The summation can be written as

$$E\left\{\hat{\mathbf{V}}_{k,m,q}[i, j]\right\} = \left[\mathbf{k}_{k,m}^{q}[i]\, \overline{\mathbf{k}}_{k,m}^{q}[j], \ldots, \mathbf{k}_{k,m}^{Q}[i]\, \overline{\mathbf{k}}_{k,m}^{Q}[j]\right] \boldsymbol{\nu}_{ij}$$

$$\boldsymbol{\nu}_{ij} = \left[\nu_{q,ij}, \cdots, \nu_{Q,ij}\right]^{\mathrm{T}}. \tag{3.32}$$

This gives us one equation per (k, m) pair in the block $\mathcal{B}_{\tilde{k},\tilde{m}}$. To stack these into a matrix, let

$$\mathcal{K}_{ij} = \begin{bmatrix} \mathbf{k}_{1,1}^{q}[i]\, \overline{\mathbf{k}}_{1,1}^{q}[j] & \cdots & \mathbf{k}_{1,1}^{Q}[i]\, \overline{\mathbf{k}}_{1,1}^{Q}[j] \\ \vdots & \ddots & \vdots \\ \mathbf{k}_{K_b,M_b}^{q}[i]\, \overline{\mathbf{k}}_{K_b,M_b}^{q}[j] & \cdots & \mathbf{k}_{K_b,M_b}^{Q}[i]\, \overline{\mathbf{k}}_{K_b,M_b}^{Q}[j] \end{bmatrix}$$

(size $K_b M_b \times Q - q + 1$) and

$$\hat{\mathbf{v}}_{ij} = \left[\hat{\mathbf{V}}_{1,1,q}[i, j], \cdots, \hat{\mathbf{V}}_{K_b,M_b,q}[i, j]\right]^{\mathrm{T}}.$$

then (3.31) becomes $E\{\hat{\mathbf{v}}_{ij}\} = \mathcal{K}_{ij}\boldsymbol{\nu}_{ij}$. If $\mathcal{K}_{ij}$ is a "tall" matrix, which requires (for $q = 1$) $K_b M_b \geq Q$, then it will be left-invertible. Applying the left inverse to both sides of the equation, we obtain

$$E\left\{(\mathcal{K}_{ij}^{\mathrm{H}}\mathcal{K}_{ij})^{-1}\mathcal{K}_{ij}^{\mathrm{H}}\hat{\mathbf{v}}_{ij}\right\} = \boldsymbol{\nu}_{ij}. \tag{3.33}$$

The least squares estimator for $\boldsymbol{\nu}_{ij}$ given $\hat{\mathbf{v}}_{ij}$ is

$$\hat{\boldsymbol{\nu}}_{ij} = (\mathcal{K}_{ij}^{\mathrm{H}}\mathcal{K}_{ij})^{-1}\mathcal{K}_{ij}^{\mathrm{H}}\hat{\mathbf{v}}_{ij} \tag{3.34}$$

which is shown by (3.33) to be unbiased. Comparing (3.30) with (3.31)–(3.32) reveals that the first element of $\hat{\boldsymbol{\nu}}_{ij}$ is the estimator we seek for $\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q}[i, j]$. Equation (3.34) is separately computed for each (i, j) to yield full matrix $\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q}$ for use in Peeling step 4.

A closer look at (3.34) reveals that it is directly related to the original peeling approach of phase centering and averaging. First define

$$\mathbf{C}_{ij} = K_b M_b (\mathcal{K}_{ij}^{\mathrm{H}}\mathcal{K}_{ij})^{-1}.$$

Table 3.1: *Ten brightest calibrator sources*

Source	Catalog Name	RA°	DEC°	SNR dB
1	3C461	350.8	58.8	-20.7
2	3C405	299.9	40.7	-21.0
3	3C86	51.8	55.3	-30.0
4	3C144	83.6	22.0	-31.6
5	3C274	187.7	12.4	-31.6
6	3C123	69.3	29.7	-31.9
7	4C+53.06	52.5	53.6	-37.2
8	4C+53.07	54.1	53.6	-37.2
9	3C33	17.2	13.3	-38.7
10	4C+55.07	53.1	55.9	-39.1

Now we can write

$$\hat{v}_{ij} = \mathbf{C}_{ij} \left[\frac{1}{K_b M_b} \mathcal{K}_{ij}^{\text{H}} \hat{\mathbf{v}}_{ij} \right].$$

Comparing this matrix equation with the summation of (3.29) reveals that the first element of the term in brackets is equal to the $(i, j)^{\text{th}}$ element of $\hat{\mathcal{V}}_{\tilde{k},\tilde{m},q}$ from (3.29). The remaining elements correspond to evaluating (3.29) for sources $q + 1$ to Q. Thus computing $\frac{1}{K_b M_b} \mathcal{K}_{ij}^{\text{H}} \hat{\mathbf{v}}_{ij}$ performs an element-wise version of the Peeling phase centering and averaging step on not just q, but for all sources q to Q. The multiplication by inversion matrix $\mathbf{C}_{ij}$ "demixes" the contributions of the sources into separate single source problems.

The estimates $\hat{\rho}_p$ in step 4 are based on all samples in the domain. The demixing algorithm works only on a single block. Therefore demixing is noisier than removing a source by conventional Peeling subtraction. The noise amplification depends on the condition of $\mathbf{C}_{ij}$. Because of the third assumption of Section 3.4.4 (large fringe rotations within a block) $\mathbf{C}_{ij}$ will be well conditioned. For large blocks, $\mathbf{C}_{ij}$ will converge to the identity matrix.

3.5   Simulation Results

3.5.1   CRB for Constant Calibration Gains and Phases

We first consider estimating calibration parameters over a small frequency-time block where they can be assumed constant. Fig. 3.3 illustrates the CRB for a realistic self calibration scenario with the full LOFAR geometry of Fig. 3.1. Station beams are pointed
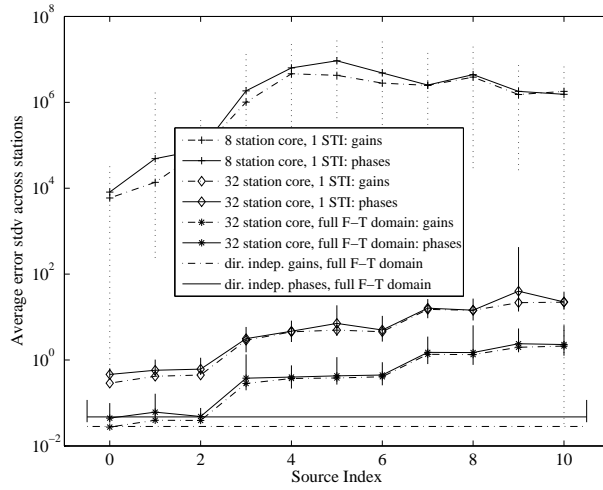
Figure 3.3: *Calibration CRB levels for constant parameters over small frequency-time span. Curves are CRB values averaged across array elements (stations) to provide a representative error level per source. Vertical bars show the range of phase errors across the 72 stations. Phase error is in radians while gain error is unitless. "Source 0" corresponds to the compact subarray which does not have source dependent calibration. Bottom horizontal curves are given as a reference for the case where calibration parameters are not source direction dependent.*

at right ascension (RA) 54.0° and declination (DEC) 55.1°.[2] An accurate model based on the existing LOFAR initial test station [30] was used for the station beam directional response, including sidelobe fine structure. For this 40 MHz observation, the -3 dB beamwidth is approximately 5° with sidelobe peak levels typically at -13 dB below the mainlobe.

The $Q = 10$ brightest radio sources after beamforming are included in the simulation. Table 3.1 lists their locations, taken from the standard 3C and 4C radio survey catalogues [31], and apparent SNRs computed from tabulated flux values. Sources 3, 7, 8, and 10 are seen within the beam mainlobe. Random "true" calibration parameters were generated using Gaussian gain magnitudes with a mean of 1.0 and standard deviation of 0.3, and phases uniformly distributed in the range $[-\pi, \pi)$.

The curves marked with a diamond in Fig. 3.3 show the CRB as function of source index for calibration on a single STI snapshot, using the central core configuration ($J_c = 32$ central core stations). The CRB is computed using equation (3.18), with constraint Jacobian $\mathbf{F}_{k,m}$ from (3.23). The 'asterisk' curves show the same for $K = 11$ frequency bins and $M = 10$ time snapshots, covering 110 STI snapshots on a sample grid with one second by 50 kHz spacing. (As with all results in Section 3.5 we assume narrowband array operation with frequency bins sufficiently narrow that no phase smearing occurs in visibility estimates. The 50 kHz bin spacing exceeds the narrowband limit, so we assume individual bins are more narrow, but selected at widely separated frequencies to reduce computational burden.) The resulting region of 10 s by 500 kHz is considered to be the maximum span that can be assumed to have constant calibration parameters. In both the single snapshot and 10 s $\times$ 500 khz cases, the use of the direction independent calibration model for the central core leads to low calibration error bounds for the first few sources. The error increases with the source index, as the source SNR decreases, and after the sixth source (third for middle single STI curves) unacceptable phase errors of more than one radian are encountered. Useful calibration for the remaining sources requires a region larger than 10 s by 500 kHz.

The top curves show the bounds for $J_c = 8$ central core stations, modeling the other 24 core stations as external stations. Since this is smaller than $Q - 1$, the calibration error becomes extremely large. The horizontal curves at the bottom are provided as a reference, and represent CRB values for the same 10 source case but where calibration parameters do *not* depend on source direction, i.e. only one complex parameter must be estimated for each station. This represents the conventional synthesis imaging problem at higher frequencies where ionospheric interaction is not strong or the aperture is smaller. The comparison illustrates the relative difficulty of direction dependent calibration, particularly for weaker sources.

---

[2]RA and DEC are astronomical polar coordinates for fixed locations in the celestial sphere used to locate deep space objects; the celestial equivalent of latitude and longitude. See e.g. http://liftoff.msfc.nasa.gov/academy/universe/radec.html.

3.5.2   CRB for Polynomial Calibration Variation

Large STI regions are needed to improve the CRB performance with direction dependent callibration. The parameters are not constant over such regions but vary smoothly, and the polynomial calibration models discussed in Section 3.4.3 will be used. To reduce the sizable computational and memory requirements, a "thinned" LOFAR array is used in all simulations to follow. Every second element from Fig. 3.1 was included, with $J = 36$ stations covering the 100 km aperture and with a central core of $J_c = 16$ stations.

In the first experiment, shown in Fig. 3.4, a basic setup without central core assumption is used, so an independent frequency-time polynomial is applied for each source-station combination. A first order in both time and frequency 2-D polynomial model was used, $\mathbf{\Gamma}(\rho, f, t) = \mathbf{Y}_1 + \mathbf{Y}_2 f + \mathbf{Y}_3 t$ and $\mathbf{\Psi}(\rho, f, t) = \mathbf{T}_1 + \mathbf{T}_2 f + \mathbf{T}_3 t$, with randomly selected "true" polynomial coefficient matrices $\mathbf{Y}_d$ and $\mathbf{T}_d$. The same 10 calibrator sources and beam steering direction as in the previous section were used in this simulation. CRB values were computed using (3.18), now with the frequency-time constraint Jacobian $\mathbf{F}_{k,m}$ from (3.26).

Fig. 3.4 presents CRB results for four linear in frequency gain and phase coefficients, i.e., entries of $\mathbf{Y}_2$ and $\mathbf{T}_2$ corresponding to source 3 and core station 2, resp. outer station 35. An important feature is that the CRBs are unacceptably high unless the estimation domain covers several seconds and/or a few hundred kilohertz. This is because sufficient frequency-time diversity due to fringe rotation in $\mathbf{K}_{k,m}$ is needed to overcome the multiple source ambiguity discussed in Section 3.4.1.

The scenario of Fig. 3.4 was repeated for Fig. 3.5 with the following changes: 1) the central core direction independent model *was* applied, by combining (3.23) and (3.26), 2) the scene contained only the first five sources from Table 3.1, and 3) the 2-D frequency-time polynomial was first order in frequency, and zero order in time: $\mathbf{\Gamma}(\rho, f, t) = \mathbf{Y}_1 + \mathbf{Y}_2 f$, $\mathbf{\Psi}(\rho, f, t) = \mathbf{T}_1 + \mathbf{T}_2 f$. This scenario will be used without change in all following experiments to exploit the central core and to reduce computational burden in simulations which involve many Monte Carlo random trials.

Comparing Figures 3.5 and 3.4(*a*) it is apparent that the central core model significantly reduces estimation error variance and the need for large time domain span. Other experiments (not shown) indicate that the reduction from 10 to five sources and use of a zero-order-in-time polynomial model were minor factors in this CRB reduction. This suggests that a self calibration algorithm should exploit the central core model if array geometry and ionospheric structure support it.

3.5.3   Peeling Calibration Performance

A full implementation of the peeling algorithm was run using synthesized array receiver data to compare its performance with the corresponding CRB. When fully operational, LOFAR will be calibrated in real time using a super computer. However, given existing computational resources and the need to run many Monte Carlo trials, the thinned array five-source scenario of Fig. 3.5 was used here.

A first order in frequency, zero order in time, 2-D polynomial was applied both for
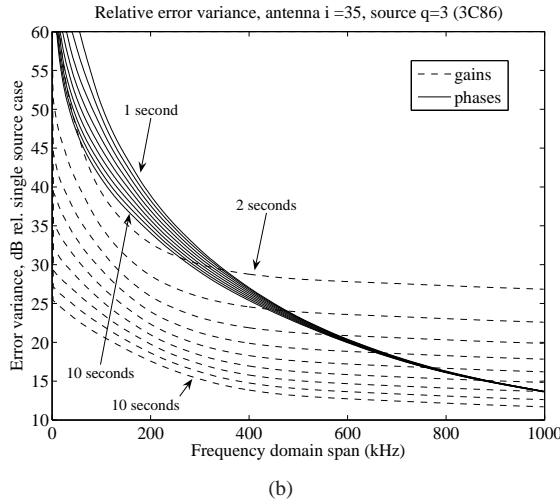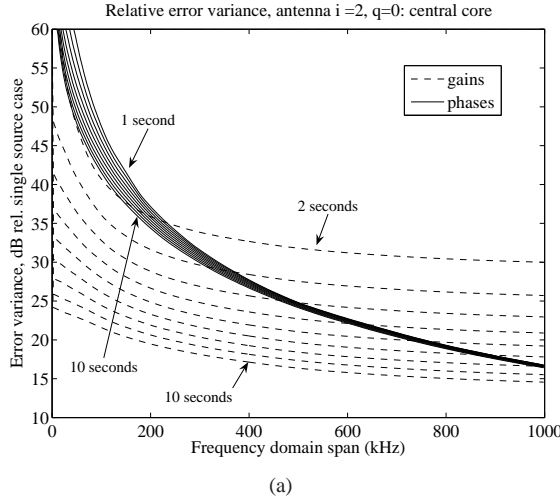
(a)

(b)

Figure 3.4: *Normalized CRB for polynomial coefficients as a function of total frequency-time span. The direction independent compact core model was* not *used. Decibel level is normalized to the single source CRB over the same frequency-time span. STI sample spacing is 1.0 second by 2.0 kHz, beginning at 40 MHz. (a) Normalized CRB for the gain, $(\mathbf{Y}_2)[2,3]$, and phase, $(\mathbf{T}_2)[2,3]$, coefficients from the linear-in-frequency polynomial term for station 2 (in the central core) and source q = 3. Curve families cover CRB dependence on time domain size from 1 to 10 seconds in 1 second increments. The 1 second gain curve is off the plot scale above. (b) CRB for $(\mathbf{Y}_2)[35,3]$ and $(\mathbf{T}_2)[35,3]$, at station 35 in an outer array arm.*
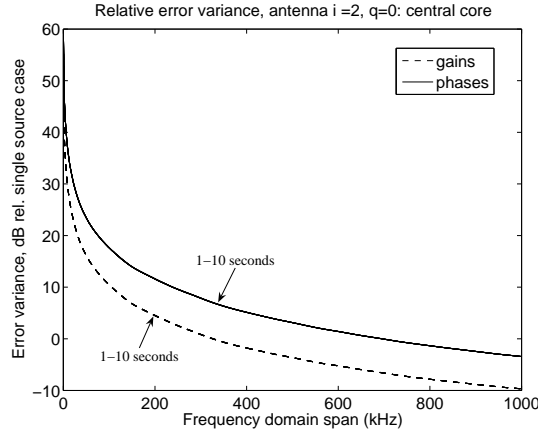
Figure 3.5: *Normalized CRB for linear in frequency coefficients* $(\mathbf{Y}_2)[2,0]$, *gain, and* $(\mathbf{T}_2)[2,0]$, *phase, when using the direction independent compact core model, corresponding to station* 2 *and "source"* $q = 0$ *in the indexing scheme of (3.22). Note that when using the compact core direction independent model, error variance on time domain span is minimal.*

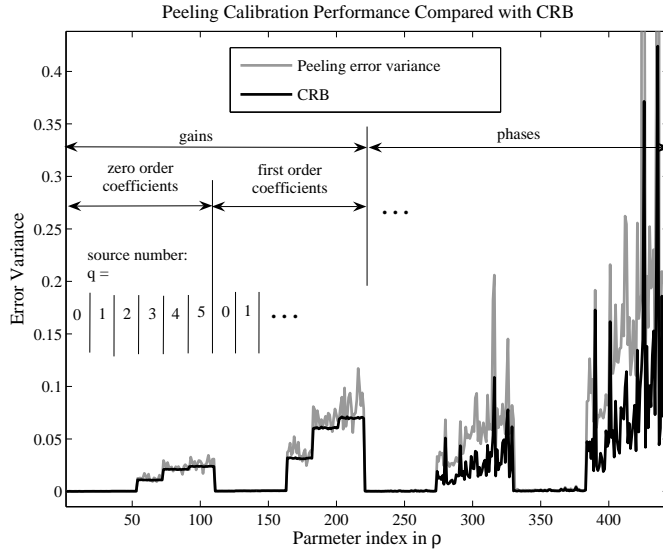**Figure 3.6:** *Peeling algorithm performance comparison with the CRB. Estimation domain size is 10 seconds by 1.0 MHz, with 1 second by 2 kHz STI sample spacing. Source index* $q = 0$ *represents the central core subarray, whose calibration polynomial coefficients do not depend on source direction. The horizontal axis parameter index for elements of* $\boldsymbol{\rho}$ *is ordered as in (3.25).*

generating the simulated array data, $\mathbf{x}_k(n)$, and in the peeling algorithm parameter model. The "true" parameter matrices were randomly generated. The frequency-time STI bin size is 2 kHz by 1 s, with $K_b = 50, M_b = 10$ the Peeling block size is 100 kHz by 10 s, and with $1 \le \tilde{k} \le 10, \tilde{m} = 1$, the total domain covers 40.0 to 41.0 MHz and 10 s. Peeling used $I = 5$ iterations.

The generated array data had a central core ($J_c = 16$), and the CRB analysis takes this into account. Thus the CRB evaluation uses the product of (3.23) and (3.26) as the constraint Jacobian, $\mathbf{F}_{k,m}$. However, the current version of Peeling does not include specific provision to exploit the central core model, so it generates distinct calibrations for each source–station combination, even though they should be direction-independent and thus identical for all sources for each central core station. As final estimate for the core array calibration parameters, only the parameters derived from the the brightest ($q = 1$) source are used.

Fig. 3.6 shows the CRB and Peeling estimation error sample variance averaged over 100 Monte Carlo trials. It is seen that Peeling closely approaches the CRB performance bound for the central core array ($q = 0$) and the two brightest sources ($q = 1, 2$). Peeling error variance is somewhat higher than the CRB for the three weaker sources ($q = 3, 4, 5$). This suggests that there is value in continued research to develop improved calibration algorithms. The plotted results are encouraging, but need to be verified on the full array and with a larger number of sources.

3.5.4 Peeling with Demixing

Performance of the combined Peeling with Demixing procedure was evaluated by computer simulation with the same models and parameter settings as used in Fig. 3.5 and Section 3.5.3. Fig. 3.7 compares the average bias error magnitude for single pass conventional Peeling with the bias from Peeling with Demixing. Estimation error variance (not shown) was acceptably low and at the same level with and without demixing. The figure shows that demixing significantly reduced bias error.

Without demixing, Peeling requires $I = 3$ passes to produce bias levels comparable to one pass of Demixed Peeling. This demonstrates the theoretical correctness of the approach described in Section 3.4.5. However, its practical utility is somewhat questionable: With our current implementation in MATLAB it takes 3.31 times as long to complete a single demix pass as to complete three passes of the regular peeling algorithm. Since both performance and complexity depend on the number of sources Q and other system parameters and assumptions, it is hard to predict how this works out in the actual LOFAR system.

3.6 Conclusions

Calibration algorithm development for LOFAR is ongoing and is critically important if the system is to achieve its ambitious scientific goals of observing the very weak signals generated during the early evolution of the universe. Interaction with the ionosphere
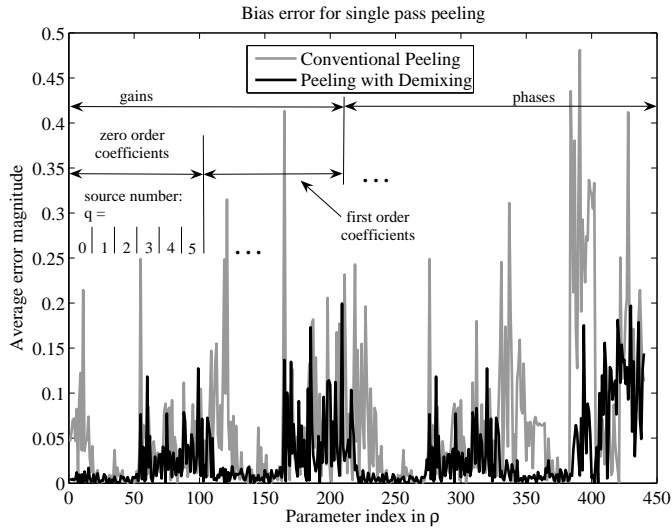
Figure 3.7: *Comparison of bias error in estimating polynomial parameter ρ_q, $0 \leq q \leq 5$ for Peeling alone, and Peeling with Demixing. One peeling pass was performed in each case. The plot illustrates the lower bias error performance of the demixing algorithm. Without demixing (not shown in this plot) three to four passes of Peeling were needed to achieve the bias levels of the lower curve. Ten Monte Carlo trials were used to compute average error magnitude. In all other details the algorithm and source parameters were identical to those used in Fig. 3.6.*

at low frequencies makes this self calibration problem significantly different and more difficult than what has been encountered in existing radio synthesis imaging instruments. It also leads to new challenges in array signal processing.

The main point of the paper was to derive and present the machinery for answering fundamental questions about calibratability for LOFAR, and in particular to assess the previously open issue of whether it is even theoretically possible to achieve self calibration. The most significant finding is that without making assumptions on the ionospheric structure, LOFAR cannot be calibrated, but with some modeling assumptions (e.g. frequency-time polynomial smoothing) and sufficient frequency-time diversity from large estimation domains, direction dependent calibration is possible. CRB analysis revealed no "show stopping" theoretical limitations on the ability to calibrate LOFAR. A central core configuration gives a significant reduction in the number of unknown parameters and thus greatly enhances the calibration performance.

The Peeling calibration algorithm was implemented and compared to the CRB using simulated data. The results indicate that at least for the limited scenarios evaluated, Peeling appears to be a viable candidate. Further algorithm development to reduce computational complexity and estimation bias due to multiple sources is warranted. Next steps will also include algorithm development to directly exploit the central core direction independent calibration model in Peeling, study of new methods to achieve more effectively reduced cross-source interference bias at the start of Peeling, and evaluation of ionospheric data and physical models to determine appropriate smoothing functions over time, frequency and space with a reduced number of parameters. To complete the picture, further studies also need to point out (*i*) the accuracy of these models (model mismatch, which translates into bias), and (*ii*) the consequences of parameter variance on the dynamic range of the image.

3.A   Appendix

Here we derive the closed form expressions for $\mathbf{M}_{k,m}$ in (3.8) which are shown in (3.9)–(3.14). Subscripts k and m are dropped for notational simplicity.

Define the Jacobians

$$\mathbf{J}\gamma_q = \frac{\partial \text{vec}(\mathbf{R})}{\partial \gamma_q^\mathrm{T}}, \quad \mathbf{J}_{\psi_q} = \frac{\partial \text{vec}(\mathbf{R})}{\partial \psi_q^\mathrm{T}}, \quad \mathbf{J}_\lambda = \frac{\partial \text{vec}(\mathbf{R})}{\partial \lambda^\mathrm{T}}\,.$$

The following expressions are useful in computing the partial derivatives:

$$\text{vec}(\mathbf{R}) = \sum_{q=1}^{Q} \sigma_q^2 \left(\overline{\mathbf{a}}_q \otimes \mathbf{a}_q\right) + \text{vec}(\mathbf{\Lambda}), \quad \frac{\partial \mathbf{a}_q}{\partial \gamma_q^\mathrm{T}} = \mathbf{\Phi}_q, \quad \frac{\partial \mathbf{a}_q}{\partial \phi_q^\mathrm{T}} = j\mathbf{E}_q$$

where $\mathbf{E}_q = \text{diag}\left(\boldsymbol{\gamma}_q \odot \mathbf{k}_q\right)$, $\boldsymbol{\Phi}_q = \text{diag}\left(\boldsymbol{\phi}_q \odot \mathbf{k}_q\right)$. The Jacobians can then be evaluated as

$$
\begin{aligned}
\mathbf{J}\boldsymbol{\gamma}_q &= \frac{\partial}{\partial \boldsymbol{\gamma}_q^{\text{T}}}\left(\sum_{p=1}^{Q}\sigma_p^2\left(\overline{\mathbf{a}}_p \otimes \mathbf{a}_p\right) + \text{vec}(\boldsymbol{\Lambda})\right) = \sigma_q^2 \frac{\partial \overline{\mathbf{a}}_q}{\partial \boldsymbol{\gamma}_q^{\text{T}}} \otimes \mathbf{a}_q + \sigma_q^2 \overline{\mathbf{a}}_q \otimes \frac{\partial \mathbf{a}_q}{\partial \boldsymbol{\gamma}_q^{\text{T}}} \\
&= \sigma_q^2 \overline{\boldsymbol{\Phi}}_q \otimes \mathbf{a}_q + \sigma_q^2 \overline{\mathbf{a}}_q \otimes \boldsymbol{\Phi}_q \\
\mathbf{J}\boldsymbol{\psi}_q &= \frac{\partial}{\partial \boldsymbol{\psi}_q^{\text{T}}}\left(\sum_{p=1}^{Q}\sigma_p^2\left(\overline{\mathbf{a}}_p \otimes \mathbf{a}_p\right) + \text{vec}(\boldsymbol{\Lambda})\right) = \sigma_q^2 \frac{\partial \overline{\mathbf{a}}_q}{\partial \boldsymbol{\psi}_q^{\text{T}}} \otimes \mathbf{a}_q + \sigma_q^2 \overline{\mathbf{a}}_q \otimes \frac{\partial \mathbf{a}_q}{\partial \boldsymbol{\psi}_q^{\text{T}}} \\
&= -j\sigma_q^2 \overline{\mathbf{E}}_q \otimes \mathbf{a}_q + \sigma_q^2 \overline{\mathbf{a}}_q \otimes j\mathbf{E}_q \\
\mathbf{J}_\lambda &= \frac{\partial \text{vec}(\boldsymbol{\Lambda})}{\partial \lambda^{\text{T}}} = \mathbf{I} \circ \mathbf{I}.
\end{aligned}
$$

An expression is derived here for $\mathbf{M}_{\boldsymbol{\gamma}_p \boldsymbol{\psi}_q}$ as given in (3.12), the other blocks are derived similarly.

$$
\begin{aligned}
\mathbf{M}_{\boldsymbol{\gamma}_p \boldsymbol{\psi}_q} &= \mathbf{J}_{\boldsymbol{\gamma}_p}^{\text{H}}(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1})\mathbf{J}_{\boldsymbol{\psi}_q} \\
&= \left(\boldsymbol{\Phi}_p \otimes \mathbf{a}_p^{\text{H}} + \mathbf{a}_p^{\text{T}} \otimes \boldsymbol{\Phi}_p^{\text{H}}\right)(\overline{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1})(-j\overline{\boldsymbol{\Phi}}_q \otimes \mathbf{a}_q + \overline{\mathbf{a}}_q \otimes -j\boldsymbol{\Phi}_q) \\
&= \sigma_p^2 \sigma_q^2 \big(-j(\boldsymbol{\Phi}_p^{\text{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{E}}_q) \otimes (\mathbf{a}_p^{\text{H}}\mathbf{R}^{-1}\mathbf{a}_q) + j(\boldsymbol{\Phi}_p^{\text{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{a}}_q) \otimes (\mathbf{a}_p^{\text{H}}\mathbf{R}^{-1}\mathbf{E}_q) \\
&\quad -j(\mathbf{a}_p^{\text{T}} \otimes \overline{\mathbf{R}}^{-1}\overline{\mathbf{E}}_q) \otimes (\boldsymbol{\Phi}_p^{\text{H}}\mathbf{R}^{-1}\mathbf{a}_q) + j(\mathbf{a}_p^{\text{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{a}}_q) \otimes (\boldsymbol{\Phi}_p^{\text{H}}\mathbf{R}^{-1}\mathbf{E}_q)\big) \\
&= 2\sigma_p^2 \sigma_q^2 \text{ Im}\left\{(\boldsymbol{\Phi}_p^{\text{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{E}}_q)(\mathbf{a}_p^{\text{H}}\mathbf{R}^{-1}\mathbf{a}_q) + (\boldsymbol{\Phi}_p^{\text{T}}\overline{\mathbf{R}}^{-1}\overline{\mathbf{a}}_q)(\mathbf{a}_p^{\text{H}}\mathbf{R}^{-1}\mathbf{E}_q)\right\}.
\end{aligned}
$$

Chapter 4

Ionospheric Modeling *

As we have seen in the previous chapter the performance of a calibration algorithm depends critically on the model for the ionospheric fluctuations. The analysis in the previous chapter only included estimation error due to noise. The analysis ignored modeling error. The variance of the estimation usually gets smaller when fewer parameters need to be estimated. On the other hand a model with more parameters can more accurately match physical reality, resulting in a smaller bias. The optimal calibration method should make a trade-off between modeling error and estimation error such that the total error is minimal.

However, a detailed model of the physical processes in the ionosphere would be impractical for the purpose of calibration. The derivation of an optimal estimator would be difficult and the computational complexity of the resulting algorithm would be too high for practical implementation. Furthermore, the current knowledge of the ionospheric processes at the scale relevant to the calibration problem at hand is incomplete. Most research on ionospheric processes deals with features down to the scale of a hundred kilometers. The smallest scale relevant for calibration is a few kilometers.

Thus we need a model that is at the same time detailed enough to capture the small scale fluctuations of the ionosphere and simple enough to allow the derivation of an algorithm of low complexity.

The usual method of describing complicated processes in a concise way is to resort to a stochastic description. For example noise is the sum of many complicated physical processes. Instead of describing these processes in detail, noise is usually described as a random (Gaussian) process.

We will use the same approach for ionospheric processes. Instead of describing the ionospheric fluctuations in full detail we will assume that the fluctuations are the result

---

of a random process. The power spectral density is assumed to be either known, or described by a model with only a few unknown parameters.

In this chapter we will present a theoretical motivation of the chosen model from literature on turbulent flow. The choice of this model is further supported by measurements from both GPS satellites and the VLA radio telescope.

4.1   Introduction

We will adopt the framework introduced by Tatarski [32] based on Kolmogorov turbulence. This framework is extensively used in (adaptive) optics where turbulence is used as a model for the refractive index fluctuations in the neutral atmosphere see Roddier [33]. Hewish [34] identified turbulence in the ionospheric F-layer as one of the possible mechanisms that could cause the small scale irregularities. Mills and Thomas [35] outlined how turbulence could be a possible mechanism capable of explaining the observed irregularities but quickly excluded it owing to the large kinematic viscosity, and hence a low Reynolds number (see below). Later, Yerg [36], with a modified approach to the kinetic theory, showed that the kinematic viscosity is smaller than the values known till then by several orders of magnitude and approaches zero in the upper ionosphere. This indicates that the Reynolds number could indeed be high enough to cause turbulence. Thompson et al. [4] mention that the use of the power law for describing the phase errors induced by the ionosphere is more realistic, and using the same exponent as used in Kolmogorov spectrum that describes the tropospheric turbulence is consistent with observations. Thus, there exist some theoretical and observational evidences for the turbulence in terrestrial ionosphere and it would seem appropriate to use the tropospheric turbulence theory to describe the plasma turbulence in the ionosphere.

4.2   Kolmogorov Turbulence

The mechanics of fluids are described by the Navier-Stokes equation,

$$\frac{\partial \mathbf{u}(\mathbf{r}, t)}{\partial t} + (\mathbf{u}(\mathbf{r}, t)\nabla)\mathbf{u}(\mathbf{r}, t) = -\frac{\nabla p(\mathbf{r}, t)}{\rho} + \nu\nabla^2\mathbf{u}(\mathbf{r}, t) \qquad (4.1)$$

where $\mathbf{u}(\mathbf{r}, t)$ is the fluid velocity, $p(\mathbf{r}, t)$ is the pressure, ρ is the density, ν is the kinematic viscosity, $\mathbf{r}$ and t are the spatial and time coordinates. The Navier-Stokes equation is non-linear because of the term $(\mathbf{u}(\mathbf{r}, t)\nabla)\mathbf{u}(\mathbf{r}, t)$. The term $\nu\nabla^2\mathbf{u}(\mathbf{r}, t)$ describes the friction between neighboring parcels of fluid. The friction causes kinetic energy to be dissipated as heat. For a stationary solution there needs to be a constant input of energy.

Let the size of the phenomenon we want to describe be of size L, and let the velocity difference be of the order U. The non-linear term is approximately U^2/L. The viscous term is $\nu U/L^2$. The ratio between the two is known as the Reynolds number, $Re = UL/\nu$. For $Re \ll 1$ the Navier-Stokes equation is linear and closed form solutions can be found in many cases. The resulting flow is laminar.

For $Re \gg 1$ stationary solutions do not exist. The non-linearity causes the solutions to be chaotic and the flow is turbulent. Consider a stable laminar flow for which $Re \ll 1$. Now start to increase the flow velocity. The Reynolds number increases also and at a certain moment the flow becomes unstable and breaks up into eddies. The eddies are smaller than the flow as a whole and they have lower Reynhold numbers. When we further increase the velocity the eddies themselves get unstable and break up into smaller eddies. For very high Reynolds numbers there exist eddies with a range of sizes from the outer scale L to the smallest scale l where viscous dissipation is dominant and kinetic energy is converted to heat. The Navier-Stokes equation for turbulent flows can only be solved numerically. A complete solution takes a lot of time, while for our purposes it is sufficient to know the statistical properties of the flow. Kolmogorov introduced some simplifying assumptions which enabled him to derive the statistical properties.

It is assumed that eddies only lose energy to eddies of smaller but comparable sizes. There is no direct flow of energy from the largest eddies to the smallest. This process has come the be known as the "Kolmogorov cascade".

In the stationary case energy can not build up at intermediate levels. Eddies of a certain size must loose the same amount of energy to the smaller eddies as they receive from larger eddies. Therefore the flow of energy can be characterized by a single number ϵ, the energy per unit time per unit mass.

It is assumed that there exists an "inertial range" between the outer scale of energy input and the inner scale where energy is dissipated. Eddies with sizes away from both the outer scale and the inner scale are not influenced by large scale effects or small scale effects. The energy contained in eddies in the inertial range only depends on ϵ. Therefore the power spectral density of the velocity variations, denoted by $\Phi_u(q)$, where q is the wavelength, must satify an equation of the form

$$(\Phi_u(q)[m^3/s^2])^x \sim (\epsilon[m^2/s^3])^y(q[m^{-1}])^z. \tag{4.2}$$

To match the the units of time (s) on both sides we need to choose $x = 3$ and $y = 2$. The last exponent z is used to match the units of length (m) by setting it to $z = -5$. This leads to

$$\Phi_u(q) \sim q^{-5/3} \tag{4.3}$$

Likewise for the three dimensional spectrum we find

$$\Phi_u(\mathbf{q}) \sim \|q\|^{-11/3} \tag{4.4}$$

This relationship has become known as Kolmogorov's law and is valid only within the inertial range $L_0^{-1} \ll q \ll l_0^{-1}$, where L_0 is the outer scale at which energy is fed into the system and l_0 is the inner scale where the energy is dissipated by viscous friction into heat.

4.2.1 Passive Conservative Additive

In [32] it is shown that the spectrum of density fluctuations of a passive and conservative additive also follows Kolmogorov's law. An additive is a substance added to the turbulent

flow. An additive is passive when it does not affect the dynamics of the flow and con-servative when it is not created nor destroyed in the flow, i.e. the density only changes due to mixing by the turbulent flow. Strictly speaking the free electrons in the ionosphere violate both conditions. They are not passive because their movements are influenced by the earth magnetic field, and not conservative because dissociation and recombination change the electron density. The resulting effect could be magnetic anisotropic density fluctuations and at times when many new electron are generated, for example at sunrise or meteor strike, a departure from the powerlaw. In our analysis we do not include these effects and assume the power spectrum of the electron density fluctuations is given by

$$\Phi_{N_e}(q_x, q_y, q_z) \sim \|\mathbf{q}\|^{-11/3}. \tag{4.5}$$

The refractive index is approximately proportional to the electron density so we assume its power spectrum is given by the same power law

$$\Phi_n(q_x, q_y, q_z) \sim \|\mathbf{q}\|^{-11/3}. \tag{4.6}$$

4.2.2   Structure function

Power law spectra can be difficult to work with since the power goes to infinity when the frequency goes to zero. At large scales a pure power law can clearly not be a valid de-scription of a physical process. The introduction of an outer scale keeps the description in the physical domain. However, if we are not interested in the large scale structures the problem can be avoided altogether by describing only local differences. The fluctua-tions of the refractive index n as function of position $\mathbf{r}$ can then described by a structure function which is defined as

$$D_n(r = \|\mathbf{r}_1 - \mathbf{r}_2\|) = \mathrm{E}\left[(n(\mathbf{r}_1) - n(\mathbf{r}_2))^2\right], \tag{4.7}$$

where $\mathrm{E}[\cdot]$ denotes the expected value. The relationship between the autocorrelation function and the structure function is given by

$$
\begin{aligned}
D_n(r) &= \mathrm{E}[(n(\mathbf{r}_1) - n(\mathbf{r}_2))^2] \\
&= \mathrm{E}[n(\mathbf{r}_1)^2 + n(\mathbf{r}_2)^2 - 2n(\mathbf{r}_1)n(\mathbf{r}_2)] \\
&= \mathrm{E}[n(\mathbf{r}_1)^2] + \mathrm{E}[n(\mathbf{r}_2)^2] - 2\kappa(\mathbf{r}_1, \mathbf{r}_2)
\end{aligned}
$$

where $\kappa(\mathbf{r}_1, \mathbf{r}_2)$ is the autocorrelation function

$$\kappa(\mathbf{r}_1, \mathbf{r}_2) = \mathrm{E}[n(\mathbf{r}_1)n(\mathbf{r}_2)]. \tag{4.8}$$

The autocorrelation function can be written as

$$\kappa(\mathbf{r}_1, \mathbf{r}_2) = 1/2(\mathrm{E}[n(\mathbf{r}_1)^2] + \mathrm{E}[n(\mathbf{r}_2)^2] - D_n(\mathbf{r}_1 - \mathbf{r}_2)). \tag{4.9}$$

The refractive index fluctuation follow a power law the structure function can be written, in the notation used by Prokhorov et al. [37], as

$$D_n(r) = C_n^2 r^{\beta-1} \tag{4.10}$$

where C_n^2 is a measure of the strength of the turbulence and $\beta = 5/3$ for Komogorov turbulence. The corresponding power spectrum is given by

$$\Phi_n(q) = A(\beta)C_n^2 q^{-\beta-2}, \tag{4.11}$$

where q is the spatial frequency and

$$A(\beta) = \Gamma(\beta + 1) \sin[(\beta - 1)\pi/2]/4\pi^2, \tag{4.12}$$

where $\Gamma(\cdot)$ is the gamma function.

4.2.3 Thin Layer Approximation

The phase change induced by a thin layer of thickness Δz is given by

$$\phi(x, y) = k \int_{H-\Delta z/2}^{H+\Delta z/2} dz n(x, y, z), \tag{4.13}$$

where $k = 2\pi/\lambda$ is the wave number. The two dimensional power spectrum is given by

$$\Phi(q_x, q_y) = 2\pi k^2 \Delta z \Phi_n(q_x, q_y, q_z = 0) \tag{4.14}$$

The corresponding structure function is

$$D_\phi(r) = 4\pi T \frac{\Gamma(1 - \beta/2)}{\beta\Gamma(1 + \beta/2)}(r/2)^\beta \tag{4.15}$$

$$= (r/s_0)^\beta \tag{4.16}$$

where $T = 2\pi k^2 \Delta z A(\beta)C_n^2$ and s_0 is the field coherence scale which is defined by $D_\phi(s_0) = 1$.

4.2.4 Propagation

After passing a turbulent layer at height H the electromagnetic wave propagates to the antennas at the ground. As long as the radius of the first Fresnel zone $r_f = \sqrt{H/k}$ is smaller than the field coherence scale the phase screen arrives at the ground essentially unaltered. For $H = 200$km and $\lambda = 4$m, $r_f = 357$m. For the datasets presented in this chapter we have found the largest value of s_0 to be 13157m and the smallest value 1645m. This is well above the Fresnel scale, so the use of geometric optics to propagate the field to the ground layer is justified.

4.2.5 Previous Measurements of Power Spectra

Measurements with the Palomar Testbed Interferometer by Linfield et al. [38] show that the power-law slope β for the neutral atmosphere was between 1.40 and 1.50 on most nights.

An overview article by Yeh and Liu [39] contains a number of references in support of power-law structure of the ionospheric fluctuations.

Velthoven [40] obtained power-law slopes above the Kolmogorov value, in the range 1.8 to 2, for wavelengths between 100km and 1000km.

4.3   GPS observations

Measurements of the signals of GPS satellites can provide valuable information on ionospheric properties. The signals from GPS satellites travel through the ionosphere and experience an ionospheric delay just like astronomical signals do. GPS satellites are actively monitored worldwide by networks of receivers. Some of these networks make their observations publicly available on the internet. In this section we describe how this data can be used to estimate the ionospheric delay by the dual frequency method and how the structure function can be estimated. The results show good agreement with a powerlaw. The slope can take a range of values, although the most frequently occurring slope is close to Kolmogorov's value.

4.3.1   Data Set

The data used in this analysis comes from the Dutch Permanent GNSS Array (DPGA). GNSS stands for Global Navigation Satellite System. The DPGA permanently observes the GPS and GLONASS satellites. Data files can be downloaded from `http://gnss1.lr.tudelft.nl/dpga/`. The highest available time resolution is one sample per second. Each hour of observation is stored in a separate file. In between the hourly measurements there are gaps of missing data of approximately 9 minutes. We have chosen for the measurements from the month January in 2006 at Cabauw, The Netherlands by the receiver identified as CAB2 (`http://gnss1.lr.tudelft.nl/dpga/station/Cabauw.html#CAB2`). The data published by the DPGA is in the compact RINEX format. The RINEX file format is the international standard for the exchange of GPS measurement data. RINEX stands for The Receiver Independent Exchange Format. The RINEX format allows users to use the same set of software to process data from different receivers. A description of the file format can be found at `http://www.ngs.noaa.gov/CORS/Rinex2.html`. To save disk space RINEX files are often compressed into the Compact RINEX format. A utility to convert Compact RINEX files to standard RINEX files can be found at `ftp://garner.ucsd.edu/pub/software/rnxcmp/source/crx2rnx.c`. The RINEX files contain among others information on the location of the receiver, the orbital parameters of the satellites and the measured carrier phase of the two carrier frequencies of the satellite.

4.3.2   Dual Frequency Method

In principle it is possible to measure the absolute TEC from satellite to receiver by measuring the time delay of a transmitted signal. In practice this is difficult because the clocks of satellite and receiver need to be synchronized and the satellite-receiver distance needs to be known up to millimeter accuracy. Furthermore, water vapor in the lower atmosphere causes an additional delay which is usually much larger than the ionospheric delay. Thus the water vapor content needs to be known too.

The problem of unknown satellite receiver distance and unknown delay in the neutral atmosphere can be overcome by using two carrier frequencies to distinguish between the

frequency dependent ionospheric delay and the other delays. The unsynchronized clocks are not a problem for estimating changes in TEC over time rather than absolute TEC.

The dual frequency method works as follows. GPS satellites have two carrier frequencies, $L1 = 1575.42$ MHz and $L2 = 1227.60$ MHz. GPS receivers track the phases of the two carriers. Because the receiver locks to an arbitrary cycle and because the oscillators of the satellite and the receiver are not synchronized, the measured phases will have an unknown offset. As long as the receiver maintains locked the offset will not change. At low SNR so called "cycle-slip" can occur. The receiver loses lock which results in a change of the offset by an integer number of cycles, or equivalently a multiple of 2π radians. It is possible to correct for an occasional cycle-slip. From here on we will assume the cycle-slips have been corrected for. Let the measured phase (in cycles) be given by $\varphi = \varphi_{true} + \varphi_{offset}$. This number can be converted to a propagation distance by multiplication with the wavelength λ_n, where n is the carrier number,

$$d(t, \lambda_n) = \varphi_n(t)\lambda_n = d_{true}(t, \lambda_n) + d_{offset,n}. \tag{4.17}$$

The measured propagation distance consists of the true, time and frequency dependent part $d_{true}(t, \lambda_n)$, and the unknown offset $d_{offset,n}$ which is different for each carrier. The true propagation distance is the sum of the geometric or physical distance and a tropospheric and an ionospheric component,

$$d_{true}(t, \lambda_n) = \varphi_n(t)\lambda_n = d_{geom}(t) + d_{trop}(t) + d_{ion}(t, \lambda_n). \tag{4.18}$$

Of these, only the ionospheric component is frequency dependent. Converting equation (2.4) for the ionospheric phase ϕ_{ion} to propagation distance d_{ion} leads to

$$d_{ion}(t, \lambda) = \frac{\lambda}{2\pi}\phi_{ion}TEC(t) = \lambda^2 C \tag{4.19}$$

where

$$C = \frac{e^2}{8\pi^2 m_e \epsilon_0 c^2} = 4.47^{-16}[\text{m}]. \tag{4.20}$$

The frequency independent components cancel by subtraction of two different measurements at two different frequencies,

$$\begin{aligned}\Delta d(t) &= d(t, \lambda_1) - d(t, \lambda_2) \\ &= d_{ion}(t, \lambda_1) - d_{ion}(t, \lambda_2) + d_{offset,1} - d_{offset,2} \\ &= TEC(t)(\lambda_1^2 - \lambda_2^2)C + d_{offset,1} - d_{offset,2}.\end{aligned}$$

The unknown offsets are still part of the equation. The offsets are constant over time so they can be canceled by subtracting two measurements at different time instants,

$$\Delta d(t_1, t_2) = \Delta d(t_1) - \Delta d(t_0) = (TEC(t_1) - TEC(t_0))(\lambda_1^2 - \lambda_2^2)C. \tag{4.21}$$

From this equation an expression for the TEC difference over time, $\Delta TEC(t_1, t_2) = TEC(t_1) - TEC(t_0)$, can be derived :

$$\Delta TEC(t_1, t_2) = \frac{\Delta d(t_1, t_2)}{(\lambda_1^2 - \lambda_2^2)C} = 9.55[\text{TECU m}^{-1}]\Delta d(t_1, t_2)[\text{m}] \tag{4.22}$$
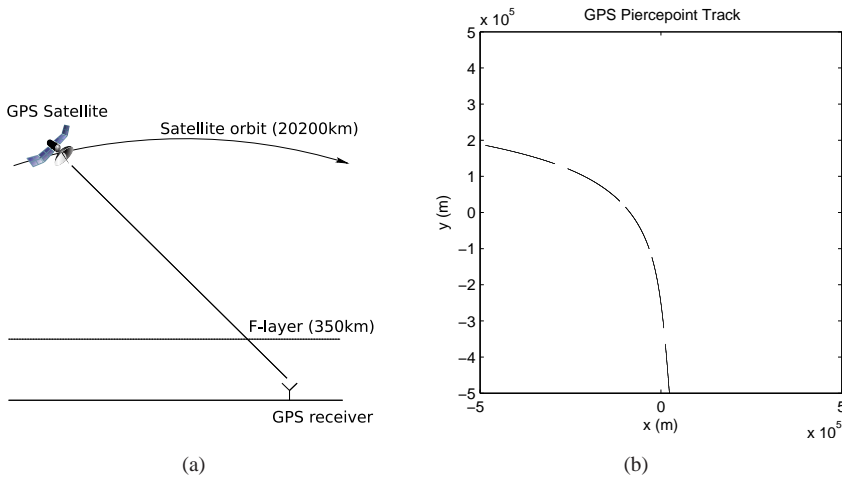
Figure 4.1: *(a) Schematic view of GPS satellite orbit, the ionospheric single layer model and the receiver. (b) Plot of the trace of the piercepoint as the GPS satellite passes over*

where the numerical value was found by substituting the wavelength of the two GPS carriers L1 and L2.

The TEC differences are rather small. To ease the interpretation of the results the TEC values have been converted to a phase shift in radians at a lower frequency. This frequency is set to 74 MHz so that a direct comparison to the 74 MHz Very Large Array (VLA) observations, presented later in this chapter, is possible.

4.3.3   GPS tracks

GPS satellites broadcast their orbital parameters which can be used to calculate the satellite positions. The parameters are stored in RINEX navigation files. These have been used to calculate the positions of the satellites. The pierce point is the intersection of the line of sight and the ionospheric layer. The ionospheric layer is assumed to be a plane, not following the curvature of the earth. The height of the layer is chosen somewhat arbitrarily to be 300km, which is approximately the height of maximum electron density. Figure 4.1(a) illustrates the model. The position of the piercepoints is given by an (x,y) coordinate pair in this plane. At low elevations the curvature of the earth can not be ignored anymore therefore an 30°elevation mask has been applied. The pierce point changes over time and traces out a track in the ionospheric plane. An example of such a track is shown in figure 4.1(b). The measured ionospheric phases together form an ionospheric phase profile.
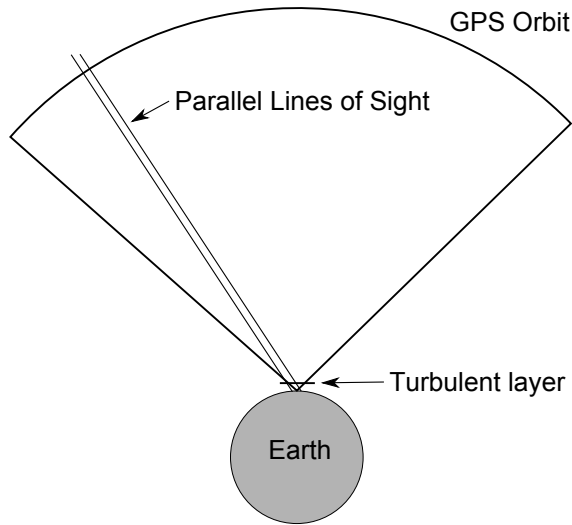
Figure 4.2: *Approximately to scale figure of the GPS orbit with a 30°elevation mask, the turbulent layer and parallel lines of sight from two antennas at 400km from each other.*

4.3.4   Detrending

The bulk ionosphere above the turbulent F-layer and the air mass effect (see below in section 4.3.5) of the mean electron density in the F-layer add a large scale trend to the ionospheric phase profile. This trend is not visible in radio astronomical observations for two reasons: 1) the lines of sight from different antennas to the same source intersects the turbulent layer at the same angle (ignoring earth curvature), 2) the width of the region of the upper ionosphere covered by the lines of sight is relatively small (not wider than the largest baseline). Figure 4.3.4 shows the GPS orbit and the lines of sight from a baseline of 400km approximately to scale. It shows the difference between probing the turbulent layer by observing the same source at different receiver locations (radio telescope) and observing a moving source from one location (GPS). The contribution of the bulk ionosphere can be approximated by a second order polynomial. To separate the turbulent fluctuations from the trend we "detrend" the data by subtracting a least squares fit of a second order polynomial. The detrending introduces some error especially at large scales because it includes part of the large scale random fluctuations in the trend and removes them. Empirically we have found that if the distance between the start and end point of a track is more than 400 km then there is only minor influence on fluctuations on the scale of 100 km and smaller. Tracks shorter than 400 km are not included in the data set.

4.3.5   Airmass Correction

After detrending the profile still needs a correction for the Airmass effect. At lower elevations the line of sight is not perpendicular to the turbulent layer. This increases the ionospheric delay by a factor $1/\cos\theta$, where θ is the zenith angle. To correct this we multiply the delays by $\cos\theta$. To calculate the zenith angle the position of the satellites needs to be known.

4.3.6   Estimation of the structure function slope

This section describes how the power law slope can be estimated from a GPS track. The model for ionospheric phase is a model over space. The observed phases are measured at different times and locations. The method described below does not distinguish between the behavior over time and over space. We assume that the ionosphere is frozen i.e. changes over space only [1].

A track consists of N samples. Each sample consists of a phase ϕ_i and a piercepoint location (x_i, y_i), where i is the sample number. There are $\frac{1}{2}N(N-1)$ different pairs of samples. For each pair we calculate the piercepoint distance $r_{i,j}$

$$r_{i,j} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{4.23}$$

The data is binned according to piercepoint distance. To obtain an even distribution of bins over the logarithm of the distance the bin sizes increase exponentially. The bin edges are given by

$$e_n = \left(\frac{d_{max}}{d_{min}}\right)^{n/K} d_{min}, \tag{4.24}$$

where d_{min} is the minimum distance, d_{max} is the maximum distance and K is the number of bins. These parameters were chosen as follows: $d_{min} = 1\,\text{km}$, $d_{max} = 100\,\text{km}$ and $K = 20$. For distances below 1 km the phase difference is small relative to the receiver noise. For distances above 100 km detrending causes an underestimation of the phase difference.

The bin mean value b_n is the average of the phase differences squared for all samples for which the piercepoint distances fall within the bin,

$$b_n = \frac{1}{M_n} \sum_{\{i,j|e_{n-1}<r_{i,j}<e_n\}} (\phi_i - \phi_j)^2, \tag{4.25}$$

where M_n is the number of pairs in bin n. In some cases, when the ionosphere is quiet, the mean value of the lower bins is below $1\,\text{rad}^2$. For these low values the measurement

---

[1] This assumption is probably not completely realistic. Indeed, the pierce point typically moves with a speed of $40-120\,\text{m/s}$ through the ionospheric plane. This is in the same order of magnitude as the wind speed at the altitude of the ionospheric layer. The typical speed of Traveling Ionospheric Disturbances (TIDs) can be several hundred meters per second. Hence we suspect that the resulting structure function will include variations over both time and space. This problem can be overcome if one has a very dense network of GPS receivers available and a method of eliminating the different offsets of each receiver. This is outside the scope of this analysis.

noise is significant, so these bins are discarded. The bin averages b_n are plotted against the bin centers $c_n = (e_{n-1} + e_n)/2$ in a double logarithmic plot, see Figure 4.3(d)

The ionospheric phase fluctuations are expected to follow a power law (4.16). When the bins are sufficiently small the bin mean values b_n are expected to be given by

$$b_n = D(c_n) = \left(\frac{c_n}{s_0} \right)^\beta. \tag{4.26}$$

Taking the logarithm of this equation yields a linear equation:

$$\log(b_n) = \log(c_n)\beta - \log(s_0). \tag{4.27}$$

After stacking the mean bin values b_n and the bin centers c_n into the vectors $\mathbf{b}$ and $\mathbf{c}$ respectively (4.27) can be written in matrix form:

$$\log \mathbf{b} = \mathbf{A}\mathbf{p}, \tag{4.28}$$

where

$$\mathbf{A} = \begin{bmatrix} \mathbf{1} & \log \mathbf{c} \end{bmatrix} \tag{4.29}$$

and

$$\mathbf{p} = \begin{bmatrix} \log s_0 \\ \beta \end{bmatrix} \tag{4.30}$$

The least squares solution to (4.28) is given by

$$\mathbf{p} = \left(\mathbf{A}^\mathsf{T}\mathbf{A} \right)^{-1} \mathbf{A}^\mathsf{T} \log \mathbf{b}, \tag{4.31}$$

The estimated power spectrum slope is p_2, the last entry of $\mathbf{p}$. A measure of the goodness of fit is the residual modeling error, given by

$$res = \left\| \left(\mathbf{I} - \mathbf{A} \left(\mathbf{A}^\mathsf{T}\mathbf{A} \right)^{-1} \mathbf{A}^\mathsf{T} \right) \log \mathbf{b} \right\|. \tag{4.32}$$

4.3.7   The procedure

The complete procedure to find the powerlaw slope from measured GPS data is summarized below.

1. Read the phases of carriers L1 and L2 from the RINEX file,

2. If necessary, correct for cycle slips,

3. Use the dual frequency method to find the ionospheric delay,

4. Detrend the ionospheric delay,

5. Calculate satellite and piercepoint positions,

6. Correct for the airmass term,

7. Estimate the structure function,

8. Fit a straight line in a double logarithmic plot.

4.3.8   Variability of the Ionosphere

The ionosphere is continuously changing. The most prominent feature on the short term is the variation over the day. During daytime radiation from the sun creates many free electrons. At night time the number of free electrons drops. On larger time scales there are seasonal differences and the eleven year solar cycle. Our data spans a period of one month, so only the variation over the day can be seen.

4.3.9   GPS Data Results and Conclusions

From the data of the month January 2006 1059 satellite tracks have been extracted. Of these 973 were longer than 400 km. For one of these tracks, starting at 18:35:46 January 3, 2006 of satellite number 24 the results of several steps in the estimation procedure are shown in figure 4.3.

Plot 4.3(a) shows the measured ionospheric phase shift at 74 MHz. The ionospheric phase shift is dominated by the contribution of the bulk ionosphere, which has been removed in figure 4.3(b). After detrending an airmass correction has been applied, the result is shown in 4.3(c). The corresponding structure function and the linear fit are shown in figure 4.3(d). This last graph shows a good agreement between the power law model and observed data for this track.

The histogram of the residual in figure 4.4 shows how well the other tracks match a power law model. The triangle marks the residual of the example track in figure 4.3. Most tracks have residuals similar to the example track. A few tracks match poorly as can be seen from the tail in the histogram. Tracks for which the residual is more than the, somewhat arbitrarily chosen, threshold of 0.18 are rejected. This removes 8% of the tracks.

The remaining 896 tracks provides a set of samples of the slope of the power law. A histogram of the slopes is shown in figure 4.5(a). The 5/3 slope predicted by Kolmogorov's law is indicated by a dashed line. The peak of the histogram is close to the theoretical value, but there is a considerable spread of slopes over the range from 1 to 2. The average slope is not constant over the day as can be seen in figure 4.5(b). A clear distinction between day and night time conditions is visible. The individual samples show that although there is a trend over the day, at any time during the day there still is a large spread over a range of slopes.

From these GPS measurements it can be concluded that most of the time a power law model is in agreement with the observed data. The most frequently found slope is close to theoretical value of 5/3, but there is is considerable spread over a range of slopes.
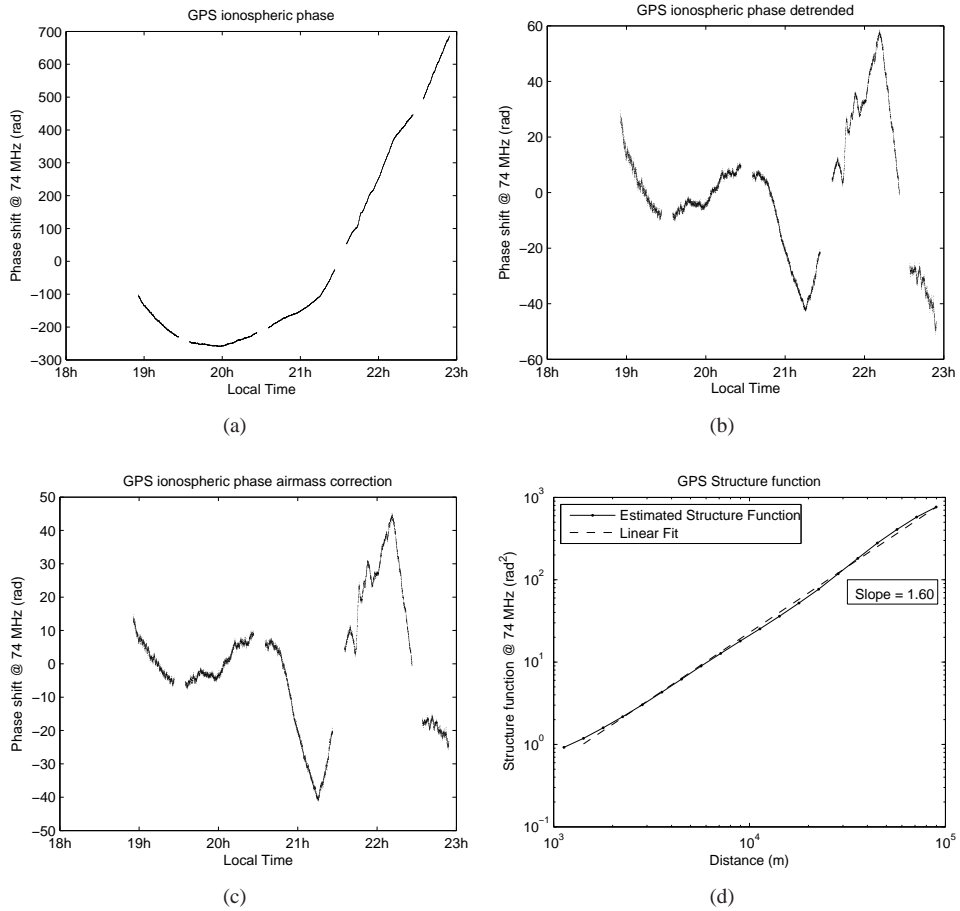
Figure 4.3: *(a) Ionospheric phase as measured by the GPS receiver. (b) Ionospheric phase after detrending. (c) Ionospheric phase after airmass correction. (d) Estimated structure function and linear fit.*
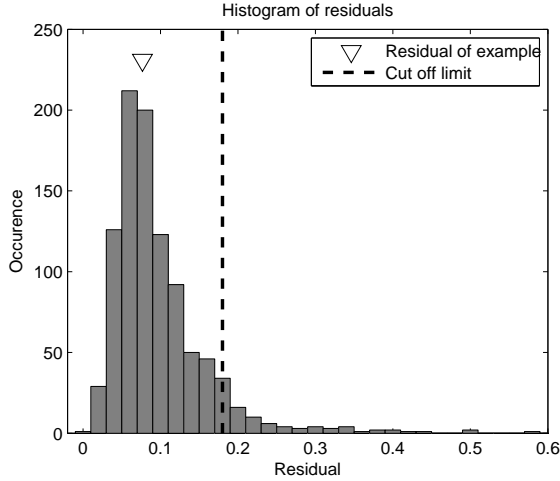
Figure 4.4: *Histogram of the residuals after fitting a power law to the data. The dashed line indicates the cut off threshold which removes 8% of the data. The triangle marks the residual of the example in figure 4.3, showing that the example has a residual that is typical for this data set.*

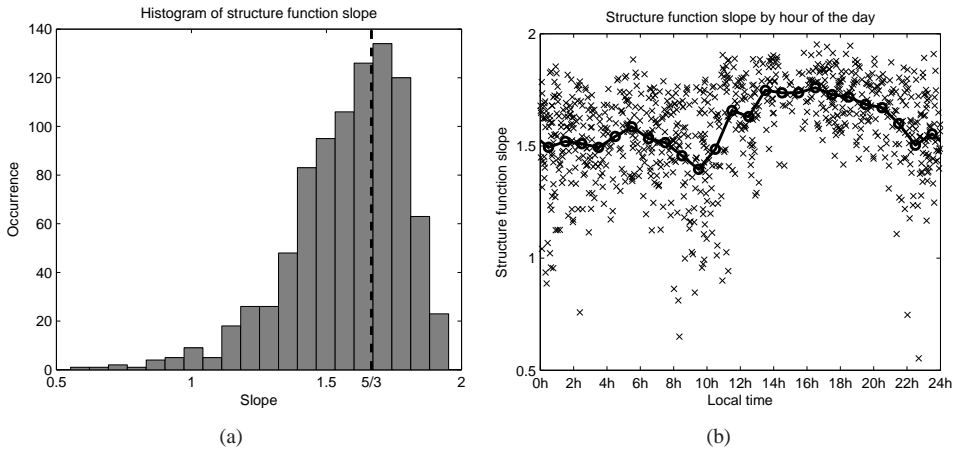(a)   (b)

Figure 4.5: *(a) Histogram of measured structure function slopes. The dashed line indicates the slope for the Kolmogorov spectrum. (b) Slope of the structure function over the day. The cross markers show the individual samples, the solid line with circular markers show the hourly averages.*

4.4   VLA observations

In this section we will analyze data obtained from the VLA Low-frequency Sky Survey (VLSS) by Cohen et al. [41]. The Very Large Array (VLA) is a radio telescope in New Mexico, U.S. wich has moveable 25 meter dishes that can be put in different configurations. The 74 MHz VLSS data has been obtained in VLA B and BnA configurations and processed using the Field Based Calibration (FBC) method developed by Cotton et al. [42]. FBC is an improvement over self-calibration accommodating a time and direction dependent phase gradient to capture the ionospheric phase fluctuations. The intricate details of the calibration procedure and the fidelity of the resulting maps can be found in [41]. Here we present a brief high level description of the data, and the calibration process that generates the intermediate data relevant for the investigations carried out in this section. This data has been analyzed before by Cohen and Röttgering [43]. They however did not fit the data to an actual model of the ionospheric phase fluctuations. Here we will fit the data to the power law model described earlier in this chapter.

4.4.1   Description of the Data

The original survey data in the VLA B configuration includes 393 distinct pointing directions in the sky. In each pointing direction, a field of view of approximately 10° is observed in three scans, each spanning approximately 25 minutes duration, separated by one or two hours. In each scan, the correlation data is integrated for 10 seconds in each baseline and stored. While processing the data, a map of the field of view is obtained in every two minutes in each scan. If the coherence time of the ionosphere is larger than 2 minutes, these maps can be considered analogous to the *short exposure images* obtained with ground based filled aperture telescopes. The sources in different directions within the field of view will be sharp, but will exhibit a random displacement from their true position or will be blurred and will exhibit a speckled intensity structure depending upon whether the ionosphere behaves like a refractive or a diffractive medium at those directions at a given time interval. The positions of all the detected sources are compared to their corresponding positions in the GHz map (NVSS catalog) and any observed differences in the positions are attributed to the random tilt induced by the ionosphere (ionospheric phase gradient) in the direction of the sources plus an unknown position shift due to a residual instrumental phase. A table of the observed position offsets of the detected sources as a function of time forms the basic data for the present study. Figure 4.6 shows a schematic representation of the data. The number of detected sources varies with time depending on the ionospheric conditions. If a source is bright enough to get detected all the time, there will be a maximum of 13 position offset values (in RA and Dec) for it in a scan.

4.4.2   Experimental Setup and Data Model

The observational setup is illustrated in figure 4.7. A field of sources is observed by an antenna array through the ionosphere. The ionosphere is assumed to be an infinitesimally
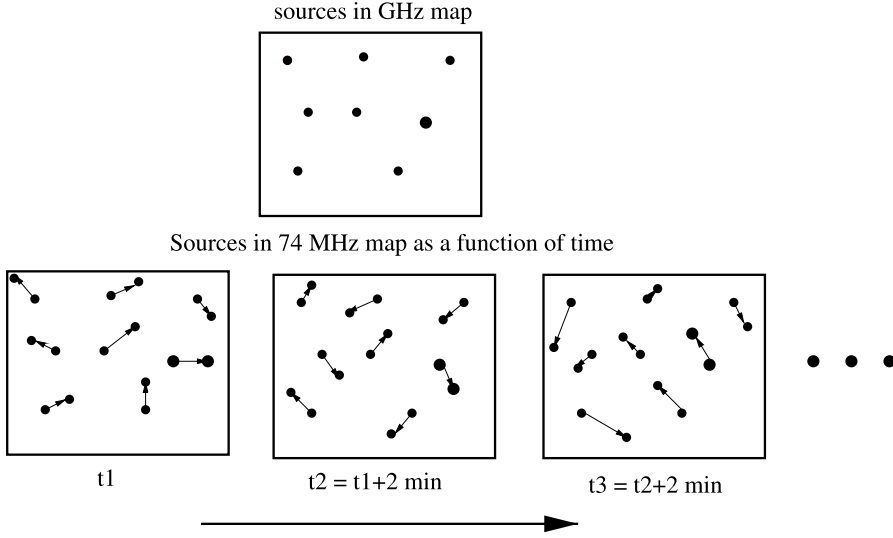
sources in GHz map

Sources in 74 MHz map as a function of time

t1   t2 = t1+2 min   t3 = t2+2 min

Figure 4.6: *Position shift of the sources with respect to the catalog as function of time due to ionospheric aberrations and residual instrumental phase*

thin layer at height H above the surface of the earth.

The array consists of M telescopes. The position of an antenna is given by the position vector $\mathbf{r} = \begin{bmatrix} x & y \end{bmatrix}^{\mathrm{T}}$. The antenna positions are collected in a ($M \times 2$) matrix

$$\mathbf{R} = \begin{bmatrix} \mathbf{r}_1^{\mathrm{T}} \\ \vdots \\ \mathbf{r}_M^{\mathrm{T}} \end{bmatrix} \tag{4.33}$$

The intersection of the path from source to antenna and the ionospheric layer is called a piercepoint. The location of the piercepoint is denoted by the length 2 vector $\mathbf{r}'$. Here we adopt the convention that whenever there are similar quantities in the ground plane and the ionospheric plane we use the same symbol, where the ionospheric quantity gets a prime ($'$). Figure 4.8 shows the two sets of piercepoints for two sources. The distance between the two sets is the piercepoint separation and it depends on both the angular distance between the sources and the height of the ionospheric layer.

For rays perpendicular to the ionospheric layer the ionospheric phase shift is given by (4.13) and will here be denoted by $\phi'(\mathbf{r}')$. For paths not perpendicular to the ionospheric layer the phase shift is scaled by the airmass term, which represents the longer distance traveled through the ionosphere. The airmass term depends on the zenith angle at the piercepoint γ' and is given by $\sec(\gamma')$. The ionospheric phase shift seen by the instrument is given by

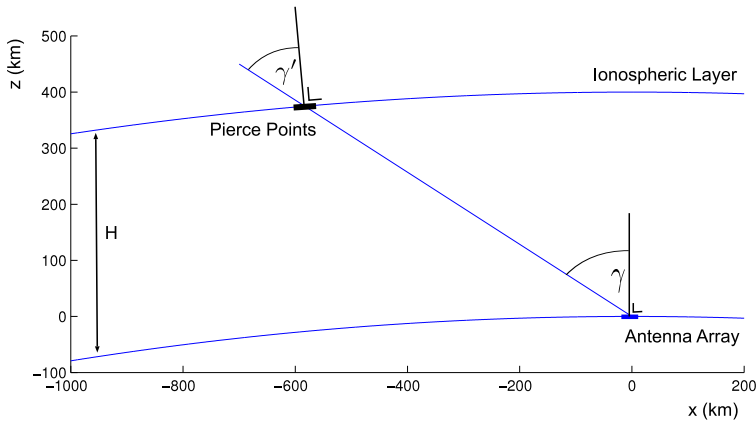$$\phi = \frac{\phi'(\mathbf{r}')}{\cos(\gamma')} \tag{4.34}$$

Figure 4.7: *Geometry of the setup*

Note that the zenith angle at the piercepoint, γ', is not equal to the zenith angle at the antenna array, γ, due to the curvature of the earth. The two are related by

$$\sin \gamma' = \frac{R}{R + H} \sin \gamma \tag{4.35}$$

where R is the radius of the earth.

4.4.3   Method

The method consist of fitting the observed position shift to the power law model of the ionospheric phase fluctuations. The observed data consists of position shifts including an unknown residual instrumental effect, while the model is for the ionospheric phase. The following processing steps are needed to do the fitting,

1. derive the differential phase gradients over the array from the observed position offsets, see Section 4.4.4,

2. derive the differential phase gradients over the array from the ionospheric phase screen, see Section 4.4.5

3. derive the statistical properties of the differential phase gradient over the array from the parameterized model for the ionospheric phase screen, see Section 4.4.6, as verification we compare the results to previous results found in literature, see Section 4.4.7,

4. fit the results of step 3 to the results of step 1 to obtain the ionospheric parameters, using the method described in Section 4.4.8.
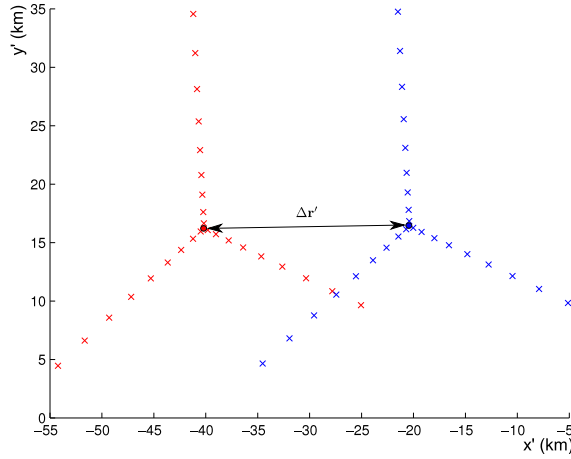
Figure 4.8: *Plot of two sets of piercepoints for two sources separated by piercepoint distance $\Delta\mathbf{r}'$. The distance depends both on the position of the sources and the height H of the layer*

4.4.4   From Position Shifts to Phase Gradients over the Array

The field based calibration method provides a set of position offsets ΔRA, Δdec as illustrated in figure 4.6. An observed position shift is caused by a phase screen over the array. We will convert the position offsets to phase shifts. The first step is convert the position offset to a change of the wave vector $\mathbf{k}$, i.e.

$$\Delta\mathbf{k} = \Delta\text{RA}\,\mathbf{k}_{\text{RA}} + \Delta\text{dec}\,\mathbf{k}_{\text{dec}}. \tag{4.36}$$

The vectors $\mathbf{k}_{\text{RA}}$ and $\mathbf{k}_{\text{dec}}$ are orthogonal to the wave vector and point in the direction of increasing right ascension and declination respectively. Over the course of the observation they slowly rotate with respect to the array. The x and y component of vector $\Delta\mathbf{k}$ can be interpreted as the phase gradient over the array. The phase gradient $\mathbf{v}$ can be written as

$$\mathbf{v} = \mathbf{I}_{2\times3}\Delta\mathbf{k} \tag{4.37}$$

where $\mathbf{I}_{2\times3}$ is the matrix

$$\mathbf{I}_{2\times3} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \tag{4.38}$$

which selects the first two entries of a length 3 vector. For a given phase gradient $\mathbf{v}$ the phases at the antennas are given by

$$\phi = \mathbf{R}\mathbf{v} \tag{4.39}$$

Due to the residual instrumental phase the observed gradients contain an unknown offset. This offset is assumed to be the same for all sources. The observed phase gradient for source i is given by

$$\mathbf{v}_i = \mathbf{v}_{\text{ion},i} + \mathbf{v}_{\text{offset}} \tag{4.40}$$

By taking differences between the gradients of two sources i and j, this unknown offset cancels and we get a quantity that is determined by the ionosphere only,

$$\Delta \mathbf{v}_{i,j} = \mathbf{v}_i - \mathbf{v}_j = \mathbf{v}_{\text{ion},i} - \mathbf{v}_{\text{ion},j}. \tag{4.41}$$

4.4.5   From Ionospheric Phases to a Gradient over the Array

In this section we will derive the phase gradient over the array as function of the ionospheric phases at the piercepoints. It is convenient to stack the ionospheric phases into a vector. For a set of pierce points $\{\mathbf{r}'_{1\ldots M}\}$ a vector of phases can be defined as

$$\phi' = \begin{bmatrix} \phi(x'_1, y'_1) \\ \vdots \\ \phi(x'_M, y'_M) \end{bmatrix}, \tag{4.42}$$

where x', y' are the coordinates of the pierce points in the ionospheric plane. The ionospheric phases are assumed to be described by a gradient. The true ionospheric phase does not follow that model so we will fit a gradient to the ionospheric phases. The gradient description and the ionospheric phase shift as seen by the array are approximately equal to each other,

$$\mathbf{PR}\mathbf{v}_{ion} \approx \frac{1}{\cos \gamma'} \mathbf{P}\phi' \tag{4.43}$$

Both sides of the equation are multiplied by the projection matrix $\mathbf{P}$ which removes the mean phase, where the mean is taken over all antennas. This mean value is invisible to the interferometer and does not cause a position shift. The projection matrix is given by

$$\mathbf{P} = \mathbf{I} - \frac{1}{M} \mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{4.44}$$

The least squares solution of (4.43) is given by

$$\mathbf{v}_{ion} = \frac{1}{\cos \gamma'} (\mathbf{PR})^{\dagger} \mathbf{P}\phi' = \mathbf{F}\phi', \tag{4.45}$$

where

$$\mathbf{F} = \frac{1}{\cos \gamma'} (\mathbf{PR})^{\dagger} \mathbf{P}. \tag{4.46}$$

The difference between two gradients $\Delta \mathbf{v}_{i,j}$ can be written in the same compact notation when the phase vectors of the both sources are stacked into one vector

$$\phi'_{i,j} = \begin{bmatrix} \phi'_i \\ \phi'_j \end{bmatrix} = \begin{bmatrix} \phi'(x'_{1,i}, y'_{1,i}) \\ \vdots \\ \phi'(x'_{M,i}, y'_{M,i}) \\ \phi'(x'_{1,j}, y'_{1,j}) \\ \vdots \\ \phi'(x'_{M,j}, y'_{M,j}) \end{bmatrix} \tag{4.47}$$

and the following matrix is defined

$$\mathbf{F}_{i,j} = \left[\begin{array}{cc} \frac{1}{\cos \gamma_i'}(\mathbf{PR})^\dagger \mathbf{P} & \frac{-1}{\cos \gamma_j'}(\mathbf{PR})^\dagger \mathbf{P} \end{array} \right] \tag{4.48}$$

The differential phase gradient can now be expressed as

$$\Delta \mathbf{v}_{i,j} = \mathbf{F}_{i,j} \phi_{i,j}' \tag{4.49}$$

From the linear expressions for the gradient (4.45) and gradient difference (4.49) in terms of the phases, the statistics of $\mathbf{v}_i$ and $\Delta \mathbf{v}_{i,j}$ can be easily derived as will be shown in the next section.

4.4.6   Gradient Statistics

In this section we will derive the statistics of the (differential) phase gradients in the form of the covariance matrices $\mathbf{C}_{\phi_i'}$ and $\mathbf{C}_{\Delta \phi_{i,j}'}$ of $\mathbf{v}_i$ and $\Delta \mathbf{v}_{i,j}$ respectively. The covariance matrix of the gradient $\mathbf{v}_i$ is defined as

$$\mathbf{C}_{\mathbf{v}_i} = \mathrm{E}\left[\mathbf{v}_i \mathbf{v}_i^{\mathrm{T}} \right]. \tag{4.50}$$

Substituting (4.45) into the equation above leads to

$$\mathbf{C}_{\mathbf{v}_i} = \mathrm{E}\left[\mathbf{F}_i \phi_i' \phi_i'^{\mathrm{T}} \mathbf{F}_i^{\mathrm{T}} \right] = \mathbf{F}_i \, \mathrm{E}\left[\phi_i' \phi_i'^{\mathrm{T}} \right] \mathbf{F}_i^{\mathrm{T}} = \mathbf{F}_i \mathbf{C}_{\phi_i'} \mathbf{F}_i^{\mathrm{T}}, \tag{4.51}$$

where $\mathbf{C}_{\phi_i'}$ is defined as

$$\mathbf{C}_{\phi_i'} = \mathrm{E}\left[\phi_i' \phi_i'^{\mathrm{T}} \right] \tag{4.52}$$

Likewise, the covariance matrix of $\Delta \mathbf{v}_{i,j}$, defined as

$$\mathbf{C}_{\Delta \mathbf{v}_{i,j}} = \mathrm{E}\left[\Delta \mathbf{v}_{i,j} \Delta \mathbf{v}_{i,j}^{\mathrm{T}} \right], \tag{4.53}$$

and substitution of (4.49) into this definition leads to

$$\mathbf{C}_{\Delta \mathbf{v}_{i,j}} = \mathrm{E}\left[\mathbf{F}_{i,j} \Delta \phi_{i,j}' \Delta \phi_{i,j}'^{\mathrm{T}} \mathbf{F}_{i,j}^{\mathrm{T}} \right] = \mathbf{F}_{i,j} \, \mathrm{E}\left[\Delta \phi_{i,j}' \Delta \phi_{i,j}'^{\mathrm{T}} \right] \mathbf{F}_{i,j}^{\mathrm{T}} = \mathbf{F}_{i,j} \mathbf{C}_{\Delta \phi_{i,j}'} \mathbf{F}_{i,j}^{\mathrm{T}}, \tag{4.54}$$

where $\mathbf{C}_{\Delta \phi_{i,j}'}$ is defined by

$$\mathbf{C}_{\Delta \phi_{i,j}'} = \mathrm{E}\left[\Delta \phi_{i,j}' \Delta \phi_{i,j}'^{\mathrm{T}} \right] \tag{4.55}$$

Now we need to find the expression for $\mathbf{C}_{\phi_i'}$ and $\mathbf{C}_{\Delta \phi_{i,j}'}$ from the power law (4.16). We assume the statistics of the ionospheric fluctuations can be described by the structure function given in equation (4.16). We further assume that the expected value of the phase is zero and that its variance is the constant $\sigma_{\phi'}^2$. From these assumption it follows that the covariance of $\phi'(\mathbf{r}_k')$ and $\phi'(\mathbf{r}_l')$ is

$$\mathrm{E}\left[\phi'(\mathbf{r}_k') \phi'(\mathbf{r}_l') \right] = \sigma_{\phi'}^2 - \frac{1}{2} D_\phi(\|\mathbf{r}_k' - \mathbf{r}_l'\|) \tag{4.56}$$

Let us now define a distance matrix $\mathbf{D}$ of which the elements are the distances between the piercepoints given by

$$d_{k,l} = \|\mathbf{r}'_k - \mathbf{r}'_l\| \tag{4.57}$$

The covariance matrix of ϕ'_i can be written as

$$\mathbf{C}_{\phi'_i} = \mathrm{E}\left[\phi'_i\phi'^{\mathrm{T}}_i\right] = -\frac{1}{2s_0^\beta}\mathbf{D}_i^{\odot\beta} + \sigma^2_{\phi'}\mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{4.58}$$

where $\mathbf{D}_i$ is the distance matrix for the piercepoints of source i, $\mathbf{D}_i^{\odot\beta}$ is $\mathbf{D}_i$ taken element wise to the power β, $\mathbf{1}$ is a vector of all ones, and the product $\mathbf{1}\mathbf{1}^{\mathrm{T}}$ is a matrix of all ones. Likewise we find for the covariance matrix of $\Delta\phi'_{i,j}$

$$\mathbf{C}_{\Delta\phi'_{i,j}} = \mathrm{E}\left[\Delta\phi'_{i,j}\Delta\phi'^{\mathrm{T}}_{i,j}\right] = -\frac{1}{2s_0^\beta}\mathbf{D}_{i,j}^{\odot\beta} + \sigma^2_{\phi'}\mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{4.59}$$

where $\mathbf{D}_{i,j}$ is the $2M \times 2M$ distance matrix for the piercepoints of sources i and j.

Now we can substitute (4.58) into (4.51) leading to

$$\mathbf{C}_{\phi'_i} = -\frac{1}{2s_0^\beta}\mathbf{D}_i^{\odot\beta} + \sigma^2_{\phi'}\mathbf{1}\mathbf{1}^{\mathrm{T}} \tag{4.60}$$

The term $\sigma^2_{\phi'}\mathbf{1}\mathbf{1}^{\mathrm{T}}$ will be projected out by multiplication with $\mathbf{F}$ so the equation simplifies to

$$\mathbf{C}_{\mathbf{v}_i} = -\frac{1}{2s_0^\beta}\mathbf{F}_i\mathbf{D}_i^{\odot\beta}\mathbf{F}_i^{\mathrm{T}} \tag{4.61}$$

Likewise for $\mathbf{C}_{\Delta\mathbf{v}_{i,j}}$ we find

$$\mathbf{C}_{\Delta\mathbf{v}_{i,j}} = -\frac{1}{2s_0^\beta}\mathbf{F}_{i,j}\mathbf{D}_{i,j}^{\odot\beta}\mathbf{F}_{i,j}^{\mathrm{T}} \tag{4.62}$$

Equations (4.61) and (4.62) are our main results which we will compare in the next section to their equivalents in literature on optical astronomy. Equation (4.62) is the model which we will fit to the data.

4.4.7   Comparison to Previous Results in Literature

The expression found by Fried [44, 45] for the variance of the norm of the position shift for a circular aperture of diameter B is given by

$$\mathrm{E}\left[\|\delta\|^2\right] = 0.357\,\lambda^2\,r_0^{-5/3}B^{-1/3}, \tag{4.63}$$

where δ is the position shift vector in radians. To compare equations (4.61) and (4.63) define the diameter B of a radio telescope as the length of the longest baseline and introduce the normalized matrices

$$\tilde{\mathbf{D}} = \mathbf{D}/B \tag{4.64}$$

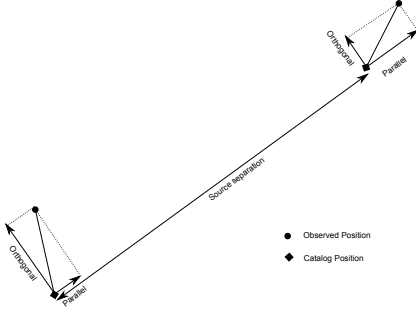$$\tilde{\mathbf{F}} = \mathbf{F}B \tag{4.65}$$

Figure 4.9: *Decomposition of the position shifts in parallel and orthogonal components with respect to the source separation*

The phase gradient can be converted to a position shift by

$$\boldsymbol{\delta} = \left[\begin{array}{c} \delta_1 \\ \delta_2 \end{array} \right] = \mathbf{v}_i/k \tag{4.66}$$

The variance of the norm of the position shifts is given by

$$\mathrm{E}\left[\|\boldsymbol{\delta}\|^2\right] = \mathrm{E}[\boldsymbol{\delta}^{\mathrm{T}}\boldsymbol{\delta}] = k^{-2}\mathrm{tr}\left(\mathbf{C}_{\mathbf{v}_i}\right) \tag{4.67}$$

Substituting $\mathbf{F}$, $\mathbf{D}$ and s_0 in equation (4.61) by the appropriately scaled equivalents $\tilde{\mathbf{F}}$, $\tilde{\mathbf{D}}$ and r_0 leads to

$$\mathrm{E}\left[\|\boldsymbol{\delta}\|^2\right] = \lambda^2 B^{\beta-2} r_0^{-\beta} \frac{1}{2\pi^2} \left(\frac{8}{\beta}\Gamma(2/\beta)\right)^{\beta/2} \mathrm{tr}(\tilde{\mathbf{F}}\tilde{\mathbf{D}}\tilde{\mathbf{F}}^{\mathrm{T}}) \tag{4.68}$$

For the VLA B configuration and $\beta = 5/3$ this becomes

$$\mathrm{E}\left[\|\boldsymbol{\delta}\|^2\right] = \lambda^2 B^{-1/3} r_0^{-5/3} \, 0.174 \, \mathrm{tr} \left[\begin{array}{cc} 1.06 & 0.0 \\ 0.0 & 1.06 \end{array} \right] \tag{4.69}$$

$$= 0.368 \lambda^2 B^{-1/3} r_0^{-5/3} \tag{4.70}$$

The 3% difference is due to the fact that equation (4.63) is based on a circular completely filled aperture, while (4.70) uses the exact antenna configuration of the VLA. Interestingly the matrix in (4.69) is a diagonal matrix with identical entries (up to the numerical accuracy). By taking the trace no information is lost. The matrix variant we have found provides not more information then Fried's scalar variant. This is only true for arrays that exhibit a large degree of symmetry, like the VLA.

Also for the differential position offset an expression has been found previously by Fried [46]. His expression is for the mean norm of the differential position offset $\Delta\boldsymbol{\delta}$ as function of the angular separation α of the sources, assuming a completely filled circular aperture. A typographical error was corrected in [47] as follows

$$\mathrm{E}\left[\|\Delta\boldsymbol{\delta}\|^2\right] = 2.91 \left(\frac{16}{\pi}\right) B^{-1/3} \int_0^\infty C_n^2(h) f_\alpha(h) dh, \tag{4.71}$$

(a) Predicted variance
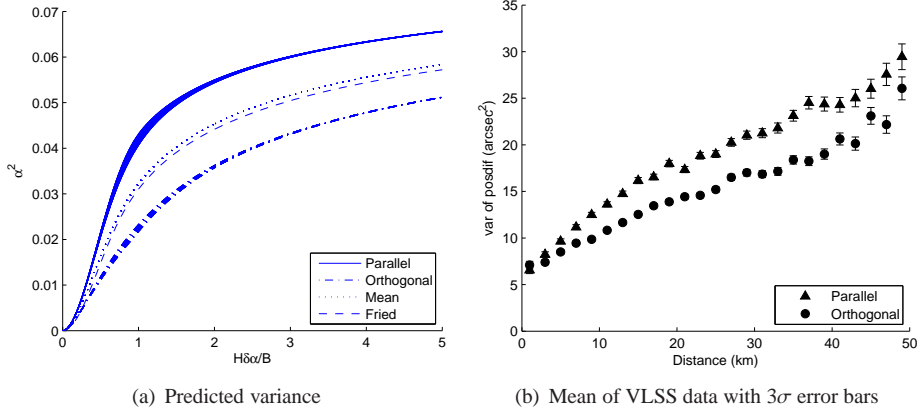
(b) Mean of VLSS data with 3σ error bars

Figure 4.10: *Variance of position shift split into a parallel and a orthogonal component with respect to the source separation. The mean of the two components matches with the expression found by Fried. The orientation of the array was changed in small steps which caused a slight broadening of the lines.*

where

$$
\begin{aligned}
f_\alpha(h) = & \\
& \int_0^{2\pi} \int_0^1 u[\arccos(u) - (3u - 2u^3)(1 - u^2)^{1/2}] \\
& \times \{0.5[u^2 + 2us\cos(\omega) + s^2]^{5/6} \\
& +0.5[u^2 - 2us\cos(\omega) + s^2]^{5/6} \\
& -u^{5/3} - s^{5/3}\}\mathrm{d}u\mathrm{d}\omega,
\end{aligned} \tag{4.72}
$$

where $s = h\alpha/B$. For the single layer model $C_n^2(h) = C_n^2\delta(h - H)$, where $\delta(\cdot)$ is Dirac's delta function, equation (4.71) reduces to

$$
\mathrm{E}\left[\|\Delta\boldsymbol{\delta}\|^2\right] = 2.91\left(\frac{16}{\pi}\right)B^{-1/3}C_n^2 f_\alpha(H) \tag{4.73}
$$

We would like to compare this result to our result (4.62). Interestingly the numerical evaluation of (4.62) is not a diagonal matrix. This means that the differential position shift is non-isotropic i.e. the shift is not equally large in all directions, but depends on the orientation with respect to the array and source separation. This anisotropy is not reflected by the equation found by Fried because it is an expression for the norm of the position shift only.

The VLA configuration exhibits a large degree of symmetry thus the orientation of the position shift with respect to the array is of minor influence. The relative orientation of the source separation and the position shift however has an appreciable effect. This
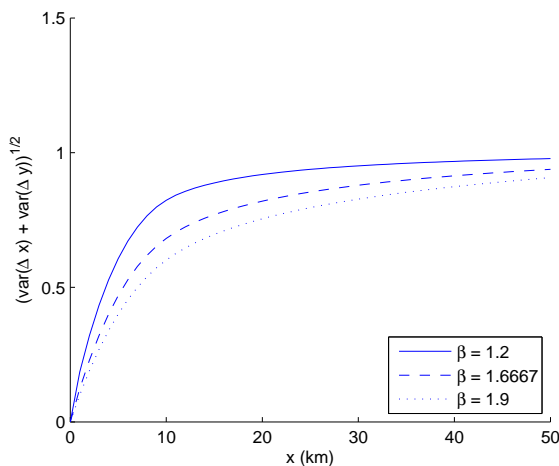
Figure 4.11: *Standard deviation of differential position offsets for configuration B for different β*

effect can be shown by decomposing the position shift in two components, one parallel, one orthogonal to the source separation as is illustrated in figure 4.9. The transformation of $\Delta\boldsymbol{\delta}$ to the vector $\Delta\tilde{\boldsymbol{\delta}}$ in the source separation oriented coordinate system can be written as a matrix multiplication

$$\Delta\tilde{\boldsymbol{\delta}} = \mathbf{U}^{\mathrm{T}}\Delta\boldsymbol{\delta}, \qquad\qquad (4.74)$$

where $\mathbf{U}$ is a 2×2 matrix of which the columns are the unit vectors $\mathbf{u}_1$ and $\mathbf{u}_2$, which are parallel and orthogonal respectively to the source separation. The correlation matrix of $\Delta\tilde{\boldsymbol{\delta}}$ is given by

$$\mathbf{C}_{\Delta\tilde{\boldsymbol{\delta}}} = \mathbf{U}^{\mathrm{T}}\mathbf{C}_{\Delta\boldsymbol{\delta}}\mathbf{U}. \qquad\qquad (4.75)$$

For the VLA in A or B configuration this matrix is practically diagonal, i.e. the transformation $\mathbf{U}$ diagonalizes $\mathbf{C}_{\Delta\boldsymbol{\delta}}$. This can be expected for an array with a large degree of symmetry and no preferred direction. The only special direction in the setup is the source separation. If there is any anisotropy in the differential position shift it will be aligned along this special direction. The results are probably different for arrays with a configuration with a preferred direction such as an East-West array or the VLA BnA configuration for which the northern arm is longer than the two other arms.

Figure 4.10 shows the two components together with the mean, and the expression found by Fried. Our results for the VLA match closely with the results by Fried for a completely filled circular aperture. The calculation has been repeated for many different orientations of the source separations with respect to the array. The results are very close to each other as can be seen in the figure where they are all plotted together, resulting in only a slight broadening of the lines.

4.4.8   Fitting Method

As we saw in the previous section the covariance of the differential phase gradients given in (4.62) depends on the height of the ionospheric layer H, the exponent in the power law β and the coherency length s_0. These unknowns will be estimated from the observed phase gradients.

An asymptotically optimal method of estimating the parameters is Maximum Likelihood (ML) estimation. In this section an algorithm will be presented to find the ML estimate.

Given a set of observations $\mathbf{x}$ with normal distribution with mean μ and variance Σ. The variance $\Sigma = \Sigma(\mathbf{p})$ is a function of parameter vector $\mathbf{p}$. The goal is to estimate $\mathbf{p}$. The ML estimator is given by the following maximization problem

$$\hat{\mathbf{p}} = \arg\max_{\mathbf{p}} \frac{1}{(2\pi)^{N/2}|\Sigma|^{1/2}} \exp(\frac{-1}{2}(\mathbf{x} - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \mu)) \tag{4.76}$$

Taking the logarithm, discarding the independent terms and reversing the sign leads to the following minimization problem

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \left[\log\det\Sigma + (\mathbf{x} - \mu)^{\mathrm{T}}\Sigma^{-1}(\mathbf{x} - \mu) \right] \tag{4.77}$$

If the covariance would be known and the mean value μ would be the unknown then the problem above would immediately be reduce to a least squares fitting problem. However in our case the mean is assumed to be zero and the covariance is the unknown parameter. This type of problems is known as covariance matching [48]. An iterative algorithm can easily be derived as is shown below.

Taking the derivative of the expression to be minimized in (4.77) and setting it to zero leads to

$$\frac{\partial \mathrm{vec}(\Sigma)^{\mathrm{T}}}{\partial \mathbf{p}} \mathrm{vec}\left(\Sigma^{-1}\left(\Sigma - \mathbf{x}\mathbf{x}^{\mathrm{T}} \right) \Sigma^{-1} \right) = \mathbf{0} \tag{4.78}$$

This problem is not easily solved but there exists a Least Squares (LS) problem that gives a result similar to (4.78). Indeed, consider

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{W}^{1/2}\mathbf{x}\mathbf{x}^{\mathrm{T}}\mathbf{W}^{1/2} - \mathbf{W}^{1/2}\Sigma\mathbf{W}^{1/2}\|_F^2. \tag{4.79}$$

Taking the derivative with respect to $\mathbf{p}$ and equating it to zero leads to

$$\frac{\partial \mathrm{vec}(\Sigma)^{\mathrm{T}}}{\partial \mathbf{p}} \mathrm{vec}\left(\mathbf{W}\left(\Sigma - \mathbf{x}\mathbf{x}^{\mathrm{T}} \right) \mathbf{W} \right) = \mathbf{0}. \tag{4.80}$$

Now if the weighting matrix $\mathbf{W}$ would have be equal to Σ^{-1}, then the problems (4.80) and (4.78) are equivalent. This leads to the following iterative method:

1. Initialize the weights $\mathbf{W} = \mathbf{I}$,

2. Solve equation (4.79),

3. If $\|\mathbf{W} - \mathbf{\Sigma}^{-1}\| > \epsilon$ then set $\mathbf{W} = \mathbf{\Sigma}^{-1}$ and go to step 2.

Here ϵ is the convergence criterion. When this iteration converges, the stationary point satisfies (4.78). If the solution of the initial least squares fit was close enough to the ML solution then the stationary point equals the ML solution.

Now as we have discussed before we expect some difficulty in estimating H and β simultaneously, because these parameters have a similar effect on the covariance. The likelihood function might not have a single peak but a range of values of H and β for which the function is close to maximum. To be able to detect this problem we will scan over a range of H and β, and estimate only s_0 by the algorithm derived above. This way we can plot the likelihood function as function of H and β.

A technicality that needs to be dealt with are the outliers. The position shifts are determined by a search for sources in an image. Detection errors lead to huge position errors that will dominate the fit when they are not removed. The removal of the outliers is a two step process. First all differential position error that exceed a fixed threshold of 200 arcsec are removed. After an initial fit the differential position shifts that exceed 3σ are rejected.

The method described above should lead to statistically optimal fits. In practice there were a few issues which limited the number of datasets for which good fits could be obtained:

- The Maximum Likelihood estimator is optimal in the case that the only sources of error are statistical fluctuations. In this case model errors are another source of significant error. The model is fairly crude; it is assumed that the ionosphere consists of a single layer of turbulence that follows a power law. The real ionosphere is distributed over a considerable height and its behavior is more complex than a simple powerlaw.

- The sampling of the piercepoints separations depends on the distances between the available calibrator sources. The most notable feature in the model for the covariance is the difference between piercepoint separations smaller and larger than the array size. A sampling that is distributed more or less evenly over both categories results in a well conditioned fitting problem. Many datasets lack a sufficient number of samples especially for the smaller piercepoint separations.

- Ideally the variance of the measurement noise should also be included as free parameter in the fitting problem. In many cases this made the fitting problem ill conditioned and visual inspection of fits showed a bad fit especially for the smaller piercepoint separations.

4.4.9   Results and Discussion

Out of 545 datasets in VLA-B configuration in the VLSS survey, three where selected to be included in this chapter. All three datasets have a relatively large number of usable samples and an even distribution over the piercepoint separation.

Figure 4.12 shows the results of the fits to the three different datasets. The plots on the left side show the logarithm of the likelihood function. Contours are plotted at the levels $\frac{-n^2}{2}$. For a linear model in Gaussian noise these contours correspond to the $n\sigma$ confidence regions. The plots on the right show the data and the model. The samples are binned according to piercepoint separation and averaged. The standard deviation within a bin is indicated by 3σ error bars. Note that the model includes the exact configuration, not just the piercepoint separation. Because of this and a selection effect, i.e. the samples that are available within a bin of piercepoint separations, even the model seems somewhat irregular in the plot.

Figure 4.12(a) shows the results for a relatively quiet ionosphere and figure 4.12(b) for a more active ionosphere. For both conditions there is a good agreement between model and observed data. That this is not always the case is shown in figure 4.12(c) where the match is not that good. The data shows a gradual increase in while the model predicts an initial steep increase which levels off for piercepoint separations larger than the size of the array.

A possible explanation would be that the electron density fluctuations are not confined to a thin layer. For different heights the characteristic turnover would occur at different angular separations. The summation over many layers results in a far more gradual transition than for a single layer. In principle one could try to fit a multilayer model to the data. There is not sufficient data to produce a meaningful result, hence a validation of the multilayer model using this data is not possible.

A distribution over heights could also explain the high value found for β. The values are very close, but not equal to 2. For $\beta = 2$ the phase screen is a random gradient with no curvature at all. The fact that the differential position shifts are not zero is solely due to the zenith angle, or airmass effect. The observed position shifts cannot be explained by the zenith angle effect alone; the fits for $\beta = 2$ are bad. However, as can be seen in figure 4.11 the closer β gets to 2 the more gradual the transition becomes, just as we would expect for a distribution over heights.

Another interesting feature is the local optimum in the lower left corner in the left panel of 4.12(c). There are quite a number datasets for which the global optimum occurred at a height below 100km and at low β. These datasets show a gradual increase without a transition or leveling off at larger piercepoint separations. The good match at low heights can be explained by realizing that a low height translates to small pierce point separations. All data points fall in the region before the transition. The model will probably fail for large piercepoint separation, but no data is available there.

It is unlikely that the position shifts are caused by the electron density fluctuations below 100km. Measurements of the electron density at these heights [49] and [50] show a typical electron density of $N_e = 10^3 \text{cm}^{-3}$. An electron density gradient over a baseline of length $B = 10$km which goes from neutral atmosphere to the typical value, and a layer of thickness $\Delta z = 10$km, leads to a position shift of

$$\frac{f_p^2}{2f^2} \frac{\Delta z}{B} = 1.5 \text{arcsec} \qquad (4.81)$$

This is an estimate of the maximum (differential) position one could expect if the turbu-

lence is localized in the lower layers. The upper bound is an order of magnitude lower than the observed position shifts.

The heights found for datasets 0700+398 and 2300+123.1 confirm the result by [40] that the strongest fluctuations are found 100km below the peak electron density which for the location of the VLA is about 300km.

4.4.10   Conclusions

Our results for the differential position shift as function of source separation extends the scalar result known in literature to a vector expression. This expression can be decomposed in two components, parallel and orthogonal. The observed data confirms that the two components behave differently.

Except for obviously wrong fits where the height is below 100km the estimated height of the ionosphere confirms earlier results stating that the strongest electron density fluctuations are found 100 km below the peak density.

The estimated values of β are higher than the 5/3 which is expected for pure Kolmogorov turbulence.
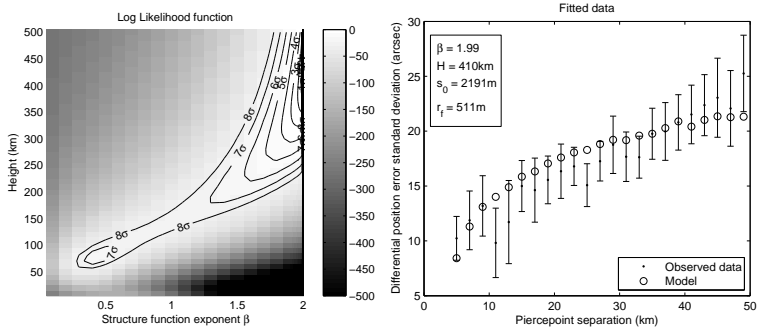
Both the high value of β and the mismatches might be explained by a multilayer or a full three dimensional model or by the presence of an inner scale. More research is needed to test this hypothesis.
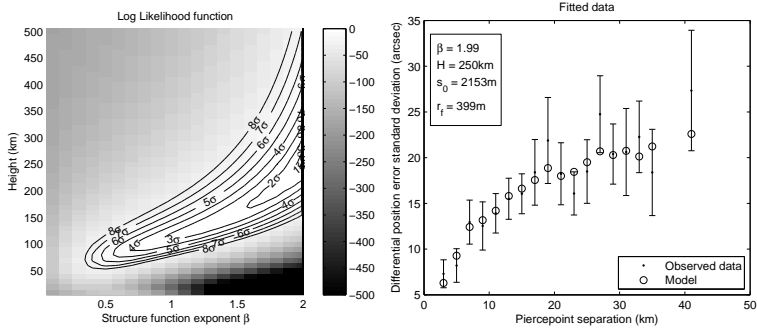
(a) Quiet Ionosphere - Field 0700+398
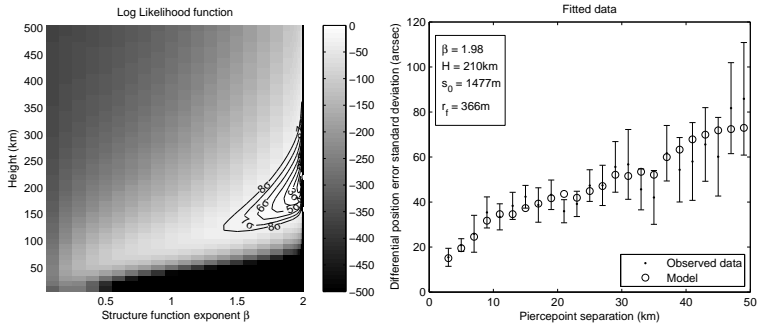
(b) Disturbed Ionosphere - Field 2300+123.1

(c) Mismatch between data and model - Field 0500+517

Figure 4.12: *Left: Loglikelihood function including contours corresponding to the $n\sigma$ confidence regions. Right: Observed data, binned including 3σ error bars and modeled data for the best fit. The fitted height H and the coherence length s_0 are shown and the Fresnel scale r_f for height H is shown too. The first two datasets show a good match between data and model for different ionospheric conditions. The last data set shows a slight mismatch. Even though most points fall within the error bars the overall trends of data and model are different.*

Chapter 5

Optimal Estimation *

One of the challenges in the design of the LOFAR radio telescope is the calibration of the ionosphere which, at low frequencies, is not uniform and can change within minutes. The number of unknown parameters quickly approaches the number of measurements and hence, structural assumptions on the ionosphere must be made, in time, frequency, and space. Using general models for the second-order statistics, we propose to use Maximum A Posteriori (MAP) estimators combined with Karhunen-Loève basis functions. The resulting estimation algorithm is shown in simulated LOFAR data to be superior to currently considered techniques. A significant advantage is that it is robust to overestimation of the number of free parameters.

5.1   Introduction

For low frequency observations (< 300 MHz) the radio astronomical community is currently developing a number of new instruments, for example the Mileura Wide Field Array (MWA) [52], the Primeval Structure Telescope (PaST) [53] and the Low Frequency Array (LOFAR) [54] which we consider in this chapter. LOFAR consists of a large number ($\sim 13,000$) of dipole antennas, arranged in 72 stations. The antennas in each station are combined to mimic a single telescope dish, which is electronically steered into the desired direction. The outputs of the stations are split into narrow frequency bins, correlated, averaged over short intervals, and stored for offline processing.

Calibration of LOFAR is essential[26], and as described in [23] has several components: calibration of the station beamshapes, and calibration of the refraction in the ionosphere. At low frequencies the effect of the ionosphere is stronger than at the higher frequencies used by most current telescopes, because the phase shift caused by the ionosphere scales with wavelength. Furthermore the beamwidths of the station beams are

---

*The results contained in this chapter have been presented at the *International Symposium on Signals, Circuits and Systems, 2007, Iasi, Romania* [51].
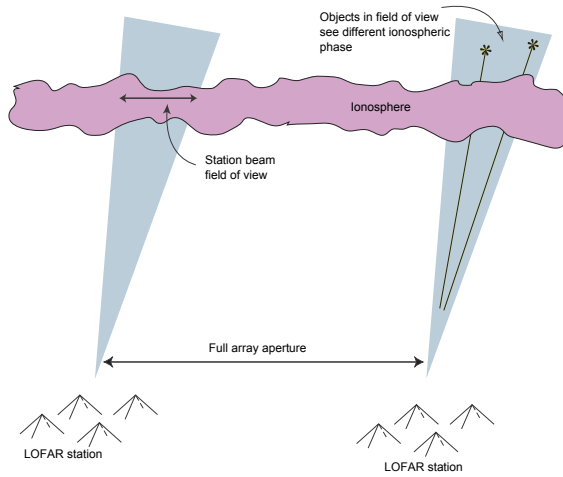
Figure 5.1: *LOFAR scenario: the ionosphere changes within the station beam and direction-dependent calibration is required. Also the ionosphere is different for different stations.*

wider than those of most existing radio telescopes. The beam projected onto the ionosphere is wider than the typical size of fluctuations within the ionosphere (in the order of a few km), see Fig. 5.1. The ionospheric phase can change considerably over the beam, therefore ionospheric calibration is direction dependent. Most stations are spaced at least a few km apart, thus it is also station dependent.

In this chapter we will describe a statistical model for ionospheric fluctuations, sufficiently simplified to be suitable for Signal Processing. The Maximum A Posteriori (MAP) estimator will be used for calibration. Simulations show that this approach is superior to the currently used method of fitting Zernike polynomials [55].

5.2 Data model and Problem Statement

5.2.1 Radio Astronomical Interferometer

A radio astronomical interferometer estimates the covariance matrices of antenna outputs by correlating them. Assume that there are M stations. Each station consists of a number of antennas, whose signals are beamformed resulting in a station signal, the equivalent of the output of a 'virtual' parabolic dish antenna. The sampled output of each station 'antenna' is split by a filter bank into narrow frequency bins. Let $\mathbf{x}_k[n]$ be a vector stacking the M station signals available at frequency k and time n. After averaging over N samples, the output of the correlator is given by

$$\hat{\mathbf{R}}_k = \frac{1}{N} \sum_{n=1}^{N} \mathbf{x}_k[n]\mathbf{x}_k[n]^{\mathrm{H}} \tag{5.1}$$

For the purpose of ionospheric calibration, we make a number of assumptions:

- Only a single calibration source is present,

- The station beamformers are pointed towards this source and the geometric delays are compensated for (the source appears at zenith),

- The instrumental phase errors are zero, and gains towards the source are unity.

These are significant simplifications that with suitable preprocessing hold true, and can be generalized later on.

The resulting data model for $\mathbf{x}_k[n]$ at the nth output sample of the kth frequency bin centered at f_k is

$$\mathbf{x}_k[n] = \mathbf{a}_k s_k[n] + \mathbf{w}_k[n] , \tag{5.2}$$

where $s_k[n]$ is the astronomical source signal, $\mathbf{w}_k[n]$ is a noise vector (i.i.d. Gaussian), and $\mathbf{a}_k$ is the spatial signature of the source given by

$$\mathbf{a}_k = \exp(i\phi_k)$$

where

$$\phi_k = C\tau f_k^{-1} \tag{5.3}$$

is a vector with M entries representing the ionospheric phases at each station (a function of frequency), and τ is a vector containing the Total Electron Content (TEC) seen by each station, which is the integral of the electron density along the line of sight towards the calibration source. The constant $C = 8422$ rad/MHz/TECU (TECU = TEC unit = 10^{16} electrons/m^2).

Under this model, the expected value of the covariance matrices is

$$\mathbf{R}_k = \mathrm{E}[\hat{\mathbf{R}}_k] = \mathbf{a}_k \mathbf{a}_k^{\mathrm{H}} \sigma_s^2 + \sigma_w^2 \mathbf{I}$$

where σ_s^2 and σ_w^2 are respectively the signal and noise power.

5.2.2 Ionospheric fluctuations

We model the ionosphere as a thin turbulent layer. The statistics of density fluctuations in a turbulent medium can be derived from Kolmogorovs theory of turbulence. Instead of using autocorrelation functions, the second-order statistics are usually given in the form of a "structure function", defined for a variable $\varphi(x)$ which is a function of a distance parameter x as

$$D_\varphi(\Delta x) = \mathrm{E}[(\varphi(x) - \varphi(x + \Delta x))^2].$$

Structure functions are used because the autocorrelation function is infinite for a pure Kolmogorov turbulent process, because all the large scale fluctuations are included. The structure function only looks at local differences which are finite.

The structure function for ionospheric TEC fluctuations over a distance r has the form [4]

$$D_{\text{TEC}}(r) = \left(\frac{f_0}{C}\right)^2 \left(\frac{r}{s_0}\right)^{\beta}$$ (5.4)

where s_0 is a reference distance and f_0 is a reference frequency. Translating this into the structure function of the phase fluctuations, we obtain

$$D_{\phi}(r, f) = \left(\frac{f_0}{f}\right)^2 \left(\frac{r}{s_0}\right)^{\beta} .$$

The expression for the structure function has the form of a power law with exponent β. The scaling is chosen such that for the reference frequency f_0 and over the reference distance s_0, the structure function $D_{\phi}(s_0, f_0) = 1$.

5.3 Calibration Algorithm

5.3.1 General Data Model

We can translate the problem into more generic terms by utilizing a general nonlinear data model of the form

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \mathbf{w}$$ (5.5)

where $\mathbf{y}$ is a vector that stacks all observations, $\boldsymbol{\theta}$ is a vector stacking the unknown parameters, and $\mathbf{w}$ is a noise vector. In this model, both the unknowns and the noise are assumed to be the result of Gaussian random processes with known covariance matrices $\mathbf{C}_{\theta}$ and $\mathbf{C}_{\mathbf{w}}$ respectively.

The generic data model is related to our application as follows. Let

$$\mathbf{y} = \text{vec}(\hat{\mathbf{R}}_k), \quad \mathbf{f}(\cdot) = \text{vec}(\mathbf{R}_k) = (\bar{\mathbf{a}}_k \otimes \mathbf{a}_k)\sigma_s^2 + \text{vec}(\mathbf{I})\sigma_w^2 ,$$

where $\text{vec}(\cdot)$ stacks the columns of a matrix into a vector, and $\otimes$ denotes the Kronecker product. The observation noise $\text{vec}(\hat{\mathbf{R}}_k - \mathbf{R}_k)$ corresponds to the noise vector $\mathbf{w}$ in (5.5). The observation noise is Wishart, not Gaussian, but for a sufficiently large N Gaussian noise is a good approximation.

The unknown parameter vector is $\boldsymbol{\tau}$, but since we cannot expect to estimate the bulk delay, we subtract the average value of $\boldsymbol{\tau}$ and define

$$\boldsymbol{\theta} = \boldsymbol{\tau} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\text{T}}\boldsymbol{\tau} .$$

The corresponding covariance matrix for $\boldsymbol{\theta}$ is, from (5.4),

$$\mathbf{C}_{\theta} = \frac{-1}{2}\left(\frac{f_0}{C}\right)^2 (\mathbf{I} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\text{T}})\left(\frac{\mathbf{D}}{s_0}\right)^{\odot\beta} (\mathbf{I} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\text{T}})$$ (5.6)

where **D** is a distance matrix containing all distances (baselines) between the antennas, and the superscript $\odot\beta$ denotes entrywise raising to the power β.

The above translates the data model into the generic model for a single frequency k and time point. It takes only the spatial structure into account. However, the model is readily generalized; the main issue is to obtain a model for the covariance matrix $\mathbf{C}_\theta$.

5.3.2   MMSE estimator

A desirable estimator is the estimator with the minimum Mean Squared Error, the MMSE, given by [29]

$$\hat{\boldsymbol{\theta}} = \arg\min_{\hat{\boldsymbol{\theta}}} \mathrm{E}[|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}|^2].$$

The solution of this minimization problem is given by

$$\hat{\boldsymbol{\theta}} = \mathrm{E}[\boldsymbol{\theta}|\mathbf{y}]$$

where the expectation is taken over the a posteriori pdf. Using Bayes' rule the a posteriori pdf is found to be

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

Both the determination of the a posteriori pdf and taking the expectation require multi-dimensional integration. In many cases an analytical solution cannot be found and numerical integration is needed. Multi-dimensional numerical integration is a computationally demanding problem, and for large problems such as LOFAR calibration this is simply not feasible.

5.3.3   MAP estimator

A good alternative is to use the Maximum A Posteriori (MAP) estimator [29],

$$
\begin{aligned}
\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{y}) = \arg\max_{\boldsymbol{\theta}} \frac{p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})}{\int p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})d\boldsymbol{\theta}} \\
&= \arg\max_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}).
\end{aligned}
$$

For the MAP no numerical integration is needed. With a Gaussian data model the MAP reduces to a Least Squares problem, because

$$p(\mathbf{y}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\mathbf{C_w}|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))^\mathrm{T}\mathbf{C_w}^{-1}(\mathbf{y} - \mathbf{f}(\boldsymbol{\theta}))\right]$$

$$p(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{k}{2}}|\mathbf{C}_\theta|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\boldsymbol{\theta}^\mathrm{T}\mathbf{C}_\theta^{-1}\boldsymbol{\theta}\right]$$

so that

$$\hat{\theta} = \arg\max_{\theta} p(\mathbf{y}|\theta)p(\theta)$$

$$= \arg\min_{\theta} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{f}(\theta))\|^2 + \|\mathbf{C}_{\theta}^{-\frac{1}{2}}\theta\|^2 . \qquad (5.7)$$

5.3.4 Interpolation

To make an image the ionospheric phase in other directions than that of the calibrators needs to be estimated. Thus some form of interpolation is required. Optimal interpolation can be achieved by including the intermediate points in the Bayesian estimation problem. The vector of unknowns

$$\theta = \left[\begin{array}{c} \kappa \\ \rho \end{array}\right]$$

now consists of two parts, κ, the phases in the direction of the calibrators, and ρ the rest of the phases. Now the MAP estimator can be applied to the extended problem

$$\hat{\theta} = \left[\begin{array}{c} \hat{\kappa} \\ \hat{\rho} \end{array}\right] = \arg\min_{\kappa,\rho} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{f}(\kappa))\|^2 + \left[\begin{array}{cc} \kappa^{\mathrm{T}} & \rho^{\mathrm{T}} \end{array}\right] \left[\begin{array}{cc} C_{\kappa\kappa} & C_{\kappa\rho} \\ C_{\rho\kappa} & C_{\rho\rho} \end{array}\right]^{-1} \left[\begin{array}{c} \kappa \\ \rho \end{array}\right]$$

The addition of more unknowns complicates the optimization problem. Fortunately the complete problem including calibration and interpolation can be separated into a distinct calibration and an interpolation problem.

$$\hat{\theta} = \arg\min_{\kappa,\rho} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{f}(\kappa))\|^2 + \left[\begin{array}{cc} \kappa^{\mathrm{T}} & \rho^{\mathrm{T}} \end{array}\right] \left[\begin{array}{cc} \mathbf{C}_{\kappa\kappa}^{-1} + \mathbf{C}_{\kappa\kappa}^{-1}\mathbf{C}_{\kappa\rho}\mathbf{S}^{-1}\mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1} & -\mathbf{C}_{\kappa\kappa}^{-1}\mathbf{C}_{\kappa\rho}\mathbf{S}^{-1} \\ \mathbf{S}^{-1}\mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1} & \mathbf{S}^{-1} \end{array}\right] \left[\begin{array}{c} \kappa \\ \rho \end{array}\right]$$

where $\mathbf{S} = C_{\rho\rho} - \mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1}\mathbf{C}_{\kappa\rho}$ is the Schur complement.

$$\hat{\theta} = \arg\min_{\kappa,\rho} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{f}(\kappa))\|^2 + \kappa^{\mathrm{T}}\mathbf{C}_{\kappa\kappa}^{-1}\kappa +$$

$$\kappa^{\mathrm{T}}\mathbf{C}_{\kappa\kappa}^{-1}\mathbf{C}_{\kappa\rho}\mathbf{S}^{-1}\mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1}\kappa - \kappa^{\mathrm{T}}\mathbf{C}_{\kappa\kappa}^{-1}\mathbf{C}_{\kappa\rho}\mathbf{S}^{-1}\rho + \rho^{\mathrm{T}}\mathbf{S}^{-1}\mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1}\kappa + \rho^{\mathrm{T}}\mathbf{S}^{-1}\rho$$

The second line of the equation can be made zero by choosing

$$\rho = \mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1}\kappa$$

The whole procedure can be summarized as follows. The first step is calibration by

$$\hat{\kappa} = \arg\min_{\kappa} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{x} - \mathbf{f}(\kappa))\|^2 + \kappa^{\mathrm{T}}\mathbf{C}_{\kappa\kappa}^{-1}\kappa.$$

The next step is interpolation by

$$\hat{\rho} = \mathbf{C}_{\rho\kappa}\mathbf{C}_{\kappa\kappa}^{-1}\hat{\kappa}.$$

This interpolation method is known as Kriging interpolation [56]

5.3.5   Karhunen-Loève transformation

In many cases the dimensionality of θ is large. The number of parameters to be estimated can be reduced by writing vector θ in terms of a sum of fewer underlying base vectors i.e.

$$\theta = \mathbf{U}_\theta \mathbf{p}$$

where $\mathbf{U}_\theta$ is a tall matrix whose columns are regarded as basis vectors. The reduced parameter vector is $\mathbf{p}$. The basis vectors in $\mathbf{U}_\theta$ can be selected in several ways:

- Data independent, e.g., by choosing polynomial functions. Zernike polynomials are often used.

- Data dependent, by computing an eigenvalue decomposition of the covariance matrix

$$\mathbf{C}_\theta = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^{\mathrm{H}} \approx \mathbf{U}_\theta \mathbf{\Lambda}_\theta \mathbf{U}_\theta^{\mathrm{H}} \tag{5.8}$$

  where $\mathbf{U}$ is a unitary matrix containing the eigenvectors, $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues. In the approximation, only the dominant eigenvectors/eigenvalues are retained in $\mathbf{U}_\theta$ and $\mathbf{\Lambda}_\theta$.

Inserting this into the MAP estimator, we obtain

$$\hat{\mathbf{p}} = \arg \min_{\mathbf{p}} \|\mathbf{C}_\mathbf{w}^{-\frac{1}{2}}(\mathbf{y} - \mathbf{f}(\mathbf{U}_\theta\hat{\mathbf{p}}))\|^2 + \|\mathbf{\Lambda}_\theta^{-\frac{1}{2}}\hat{\mathbf{p}}\|^2 . \tag{5.9}$$

After estimating $\hat{\mathbf{p}}$, an estimate of $\hat{\theta}$ is obtained as $\hat{\theta} = \mathbf{U}_\theta\hat{\mathbf{p}}$.

5.3.6   Unknown hyperparameters

In the context of Bayesian estimation, the parameters which parameterize the a priori distribution are called the hyperparameters. In our case the hyperparameters are β and s_0. If they are unknown they need to be estimated too. The MAP estimator can easily be extended to incorporate this by simply extending the search space with the extra unknowns. However, this makes the problem much harder because it changes from a large least squares problem, to a large generic non-linear problem. A solution is to alternately estimate θ, using least squares, and the other two parameters using a generic non-linear solver:

1. Initialize β and s_0 with some reasonable guess

2. Estimate $\mathbf{p}$ using (5.9)

3. Estimate β and s_0 using a non-linear solver as

$$\arg \max_{\beta,s_0} \frac{1}{(2\pi)^{\frac{k}{2}}|\mathbf{C}_\theta(\beta, s_0)|^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\theta^{\mathrm{T}}\mathbf{C}_\theta^{-1}(\beta, s_0)\theta\right]$$

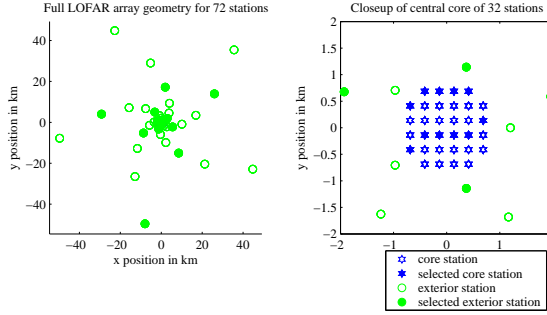4. Check for convergence of β and s_0, if necessary go to step 2.

Figure 5.2: *Station configuration used for simulation. Only the selected stations were actually included in the simulation.*

5.4 Simulations

In this section, we demonstrate the applicability of the proposed method to the LOFAR calibration problem. Several simplifications were made to allow for sufficient Monte Carlo runs. The ionosphere is assumed to be a thin layer above the array. The TEC values of the ionospheric layer are a function of position. They are assumed to be the result of a random process with a Kolmogorov spectrum. The TEC values of interest are the ones at the pierce-points where the line of sight intersects the ionospheric layer. Ignoring the curvature of the earth, the distance between the pierce-points equals the distance between the stations. Figure 5.2 shows a configuration of 72 stations which is similar to the actual LOFAR configuration. One third of the stations was selected to be included in the simulation, so the number of antennas $M = 24$. Let $\mathbf{v}_i$ be the vector describing the position of the ith station. The entries of the distance matrix $\mathbf{D}$ are given by $d_{i,j} = \|\mathbf{v}_i - \mathbf{v}_j\|$. From the distance matrix the covariance matrix $\mathbf{C}_\theta$ can be found using (5.6). The parameters used to generate the data are $\beta = 5/3$, $f_0 = 100$ MHz and $s_0 = 3000$ m, i.e. a pure Kolmogorov spectrum with an r.m.s phase fluctuation of 1 radian over a distance of 3000 m at 100 MHz.

The relative TEC values $\boldsymbol{\tau}$ are generated as

$$\boldsymbol{\theta} = \mathbf{C}_\theta^{1/2}\mathbf{w}\,,$$

where $\mathbf{w}$ is zero mean i.i.d. Gaussian noise. The resulting $\boldsymbol{\theta}$ is jointly Gaussian with covariance $\mathbf{C}_\theta$. The TEC values are subsequently used to construct the spatial signatures $\mathbf{a}_k$. Data samples $\mathbf{x}_k[n]$ are generated using Gaussian random signals $s_k[n]$ and noise $\mathbf{w}_k[n]$ according to equation (5.2). The covariance estimates are then obtained from (5.1).

For the simulation we have used 501 frequency bins of 1 kHz with center frequencies ranging from 100 Mhz to 100.5 MHz. The integration time is 1 second so each covariance estimate is based on $N = 1000$ samples. The signal to noise ratio σ_s^2/σ_w^2 was -30dB. For each Monte Carlo run a new set of TEC values and covariance data was generated.
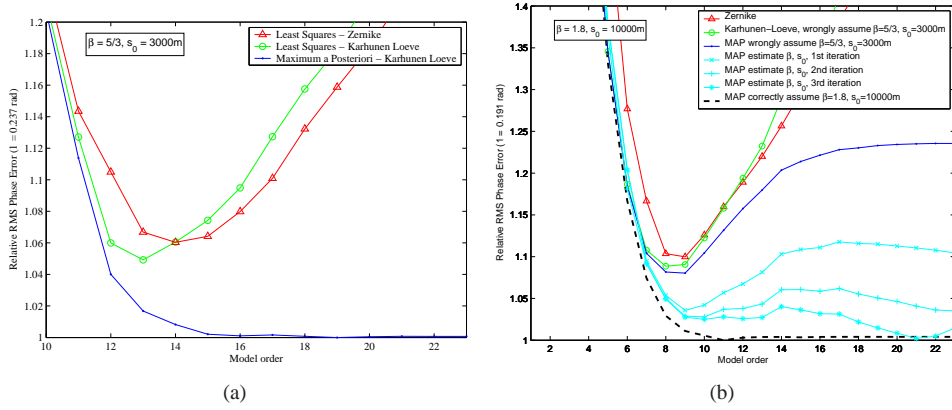
Figure 5.3: *Estimation performance as function of model order selection: (a) known model parameters, (b) estimated model parameters. The error is scaled relative to the minimum in each plot.*

Based on the simulated data the TEC values are estimated using three different methods: Least Squares using Zernike polynomials, Least Squares using a Karhunen-Loève basis, and the MAP using a Karhunen-Loève basis. All methods are based on equation (5.9). When we omit the term $\|\mathbf{\Lambda}_\theta^{-\frac{1}{2}}\hat{\mathbf{p}}\|$, the method reduces to an ordinary least squares fit. The basis $\mathbf{U}_\theta$ consists either of Zernike polynomials or the Karhunen-Loève basis computed from (5.8). When the term is included, the method is a (truncated) MAP. In each case, the size of the basis, or the model order, can be varied from 1 to $M - 1$. The maximum order is one less than the number of antennas because only the relative TEC τ is estimable.

From the estimated τ, the ionospheric phases ϕ_k are computed as in (5.3). The error measure is the r.m.s. phase error at the reference frequency $f_0 = 100$MHz.

In figure 5.3(a)(*a*) the error is plotted against the selected model order. It is seen that increasing the order reduces the model error, but at the same time the estimates get noisier. Initially incrementing the order will result in a lower total error, but at some point the additional noise outweighs the reduction of the modeling error.

Using the optimal basis (5.8) improves the performance. The lowest error of the LS method using the Karhunen-Loève basis is below the best performance of the Zernike polynomials, and is also reached at a lower order. However, to reach the lowest attainable error of the Least Squares method one needs to know at what order the optimum is reached.

The MAP estimator is not only always better than the Least Squares methods, but it is also guaranteed that the performance will improve with increasing model order. The fact that the algorithm is robust to overestimation of the number of free parameters is a significant advantage.

The increased performance is the result of exploiting the *a priori* information. Of

course, if the assumed prior does not match the actual distribution, the performance suffers, as can be seen in figure 5.3(b), where the MAP estimation is done using wrong settings for the hyperparameters (β, s_0). The performance is greatly improved by estimating the hyperparameters iteratively, as proposed in section 5.3.6.

In the next chapter the estimation method described in this chapter will be applied to actual observations from the Very Large Array (VLA).

Application of the MAP estimator to 74 MHz VLA Data *

This chapter presents a description and first results of SPAM (Source Peeling and Atmospheric Modeling), a new calibration method that attempts to iteratively solve and correct for ionospheric phase errors. SPAM has been developed by H.T. Intema at Leiden Observatory, Leiden University. The method uses a variant of the peeling technique, which we analyzed in Chapter 3. SPAM proved to be ideal to integrate the ionspheric model presented in chapter 4 and the MAP estimator presented in chapter 5 and test these on real observed data. In cooperation parts of the techniques described in Chaper 5 were included in SPAM.

To model the ionosphere SPAM uses a time-variant, 2-dimensional phase screen at fixed height above the Earth's surface. Spatial variations are described by a truncated set of discrete Karhunen-Loève base functions, optimized for an assumed power-law spectral density of free electrons density fluctuations described in Chapter 4, and a given configuration of calibrator sources and antenna locations. The model is constrained using antenna-based gain phases from individual self-calibrations on the available bright sources in the field-of-view. Application of SPAM on three test cases, a simulated visibility data set and two selected 74 MHz VLA data sets, yields significant improvements in image background noise (5–75 percent reduction) and source peak fluxes (up to 25 percent increase) as compared to the existing self-calibration and field-based calibration

---

*The contents of this chapter have been accepted for publication as a paper by H. T. Intema, S. van der Tol, W. D. Cotton, A. S. Cohen, I. M. van Bemmel and H. J. A. Röttgering entitled "Ionospheric Calibration of Low Frequency Radio Interferometric Observations using the Peeling Scheme: I. Method Description and First Results" in *Astronomy and Astrophysics* The main text of this paper was written almost entirely by Intema. To the paper a brief description of the MAP estimator was added as an appendix. That part is omitted here since a more detailed description can be found in chapter 5 of this thesis. However, the implementation of the estimator in software has been done in close cooperation with the author of this thesis. Intema has agreed to include the almost verbatim text of the paper in this thesis.

methods, which indicates a significant improvement in ionospheric phase calibration accuracy.

6.1   Introduction

Radio waves of cosmic origin are influenced by the Earth's atmosphere before detection at ground level. At low frequencies (LF; $\lesssim 300$ MHz), the dominant effects are refraction, propagation delay and Faraday rotation caused by the ionosphere [e.g. 4]. For a ground-based interferometer (array from here on) observing a LF cosmic source, the ionosphere is the main source of phase errors in the visibilities. Amplitude errors may also arise under severe ionospheric conditions due to diffraction or focussing [e.g. 57].

The ionosphere causes propagation delay differences between array elements, resulting in phase errors in the visibilities. The delay per array element (antenna from here on) depends on the line-of-sight (LoS) through the ionosphere, and therefore on antenna position and viewing direction. The calibration of LF observations requires phase corrections that vary over the field-of-view (FoV) of each antenna. Calibration methods that determine just one phase correction for the full FoV of each antenna (like self-calibration; [e.g. 11]) are therefore insufficient.

Ionospheric effects on LF interferometric observations have usually been ignored for several reasons: (i) the resolution and sensitivity of the existing arrays were generally too poor to be affected, (ii) existing calibration algorithms (e.g., self-calibration) appeared to give reasonable results most of the time, and (iii) a lack of computing power made the needed calculations prohibitly expensive. During the last 15 years, two large and more sensitive LF arrays have become operational: the VLA at 74 MHz [58] and the GMRT at 153 and 235 MHz [59]. Observations with these arrays have demonstrated that ionospheric phase errors are one of the main limiting factors for reaching the theoretical image noise level.

For optimal performance of these and future large arrays with LF capabilities (such as LOFAR, LWA and SKA), it is crucial to use calibration algorithms that can properly model and remove ionospheric contributions from the visibilities. Field-based calibration [42] is the single existing ionospheric calibration & imaging method that incorporates direction-dependent phase calibration. This technique has been succesfully applied to many VLA 74 MHz data sets, but is limited by design for use with relatively compact arrays.

In Section 6.2, we discuss ionospheric calibration in more detail. In Section 6.3, we present a detailed description of SPAM, a new ionospheric calibration method that is applicable to LF observations with relatively large arrays. In Section 6.4, we present the first results of SPAM calibration on simulated and real VLA 74 MHz observations and compare these with results from self-calibration and field-based calibration. Conclusions and a discussion are presented in Section 6.5.

6.2 Ionosphere and Calibration

In this Section, we describe some physical properties of the ionosphere, the phase effects on radio interferometric observations and requirements for ionospheric phase calibration.

6.2.1 The Ionosphere

The ionosphere is a partially ionised layer of gas between $\sim$ 50 and 1000 km altitude over the Earth's surface [e.g. 60]. It is a dynamic, inhomogeneous medium, with electron density varying as a function of position and time. The state of ionization is mainly influenced by the Sun through photo-ionization at UV and short X-ray wavelengths and through injection of charged particles from the solar wind. Ionization during the day is balanced by recombination at night. The peak of the free electron density is located at a height around 300 km. The free electron column density along a LoS through the ionosphere is generally referred to as *total electron content*, or TEC. The TEC unit (TECU) is 10^{16} m^{-2} which is a typically observed value at zenith during nighttime.

The refraction and propagation delay are caused by a varying refractive index n of the ionospheric plasma along the wave trajectory. For a cold, collisionless plasma without magnetic field, n is a function of the free electron density n_e and is defined by [e.g. 4]

$$n^2 = 1 - \frac{\nu_p^2}{\nu^2},$$ (6.1)

with ν the radio frequency and ν_p the plasma frequency, given by

$$\nu_p = \frac{e}{2\pi} \sqrt{\frac{n_e}{\epsilon_0 m}},$$ (6.2)

with e the electron charge, m the electron mass, ϵ_0 the vacuum permittivity. Typically, for the ionosphere, ν_p ranges from 1–10 MHz, but may locally rise up to $\sim$ 200 MHz in the presence of sporadic E-layers (clouds of unusually high free electron density). Cosmic radio waves with frequencies below the plasma frequency are reflected by the ionosphere and do not reach the Earth's surface. For higher frequencies, the spatial variations in electron density cause local refractions of the wave (Snell's Law) as it travels through the ionosphere, thereby modifying the wave's trajectory. The total propagation delay, integrated along the LoS, results in a phase rotation given by

$$\phi^{\text{ion}} = -\frac{2\pi\nu}{c} \int (n-1)\, dl,$$ (6.3)

with c the speed of light in vacuum. For frequencies $\nu \gg \nu_p$, this can be approximated by

$$\phi^{\text{ion}} \approx \frac{\pi}{c\nu} \int \nu_p^2 dl = \frac{e^2}{4\pi\epsilon_0 mc\nu} \int n_e\, dl.$$ (6.4)

where the integral over n_e on the right is the TEC along the LoS. Note that this integral depends on the wave's trajectory, and therefore on local refraction. Because the refractive index is frequency-dependent, the wave's trajectory changes with frequency. As a consequence, the apparent scaling relation $\phi^{\text{ion}} \propto \nu^{-1}$ from Equation 6.4 is only valid to first order in frequency.

Although bulk changes in the large scale TEC (e.g., a factor of 10 increase during sunrise) have the largest amplitudes, the fluctuations on relatively small spatial scales and short temporal scales are most troublesome for LF interferometric observations. Most prominent are the traveling ionospheric disturbances (TIDs), a response to acoustic-gravity waves in the neutral atmosphere [e.g. 40]. Typically, medium-scale TIDs are observed at heights between 200 and 400 km, have wavelengths between 250 and 400 km, travel with near-horizontal velocities between 300 and 700 km h^{-1} in any direction and cause 1–5 percent variations in TEC [4].

The physics behind fluctuations on the shortest spatial and temporal scales is less well understood. Temporal and spatial behaviour may be coupled through quasi-frozen patterns that move over the area of interest with a certain velocity and direction [57]. Typical variations in TEC are on the order of 0.1 percent, observed on spatial scales of tens of kilometers down to a few km, and time scales of minutes down to a few tens of seconds. The statistical behaviour of radio waves passing through this medium suggests the presence of a turbulent layer with a power-law spectral density of free electron density fluctuations $P_{n_e}(q) \propto q^{-\alpha}$ [e.g. 4], with $q \equiv |\vec{q}|$ the magnitude of the 3-dimensional spatial frequency. $P_{n_e}(q)$ is defined in units of electron density squared per spatial frequency. The related 2-dimensional structure function of the phase rotation ϕ of emerging radio waves from a turbulent ionospheric layer is given by

$$D_\phi = \langle [\phi(\vec{x}) - \phi(\vec{x} + \vec{r})]^2 \rangle \propto r^\gamma, \tag{6.5}$$

where $\vec{x}$ and $\vec{x} + \vec{r}$ are Earth positions, $r \equiv |\vec{r}|$ is the horizontal distance between these two points, $\langle \ldots \rangle$ denotes the expected value and $\gamma = \alpha - 2$. For pure Kolmogorov turbulence, $\alpha = 11/3$, therefore $\gamma = 5/3$.

Using differential Doppler-shift measurements of satellite signals, van Velthoven [40] found a power-law relation between spectral amplitude of small-scale ionospheric fluctuations and latitudinal wave-number with exponent $\alpha/2 = 3/2$. Combining with radio interferometric observations of apparent cosmic source shifts, van Velthoven derived a mean height for the ionospheric perturbations of 200–250 km. Through analysis of differential apparent movement of pairs of cosmic sources in the VLSS, Cohen and Röttgering [43] find typical values for $\gamma/2$ of 0.50 during nighttime and 0.69 during daytime. Direct measurement of phase structure functions from different GPS satellites (van der Tol, *unpublished*) shows a wide distribution of values for γ that peaks at ~ 1.5. On average, these results indicate the presence of a turbulent layer below the peak in the free electron density that has more power in the smaller scale fluctuations than in the case of pure Kolmogorov turbulence. Note that for individual observing times and locations, the behaviour of small-scale ionospheric fluctuations may differ significantly from this average.

6.2.2 Image Plane Effects

Interferometry uses the phase differences as measured on baselines to determine the angle of incident waves, and is therefore only sensitive to TEC differences. A baseline is sensitive to TEC fluctuations with linear sizes that are comparable to or smaller than the baseline length. At 75 MHz, a 0.01 TECU difference on a baseline causes a ∼ 1 radian visibility phase error (Equation 6.4). Because the observed TEC varies with time, antenna position and viewing direction, visibility phases are distorted by time-varying differential ionospheric phase rotations.

An instantaneous spatial phase gradient over the array in the direction of a source causes an apparent position shift in the image plane [e.g. 43], but no source deformation. If the spatial phase behaviour deviates from a gradient, this will also distort the apparent shape of the source. Combining visibilities with different time labels while imaging causes the image plane effects to be time-averaged. A non-zero time average of the phase gradient results in a source shift in the final image. Both a zero-mean time variable phase gradient and higher order phase effects cause smearing and deformation of the source image, and consequently a reduction of the source peak flux (see [55] for an example). In the latter case, if the combined phase errors behave like Gaussian random variables, a point source in the resulting image experiences an increase of the source width and reduction of the source peak flux, but the total flux (the integral under the source shape) is conserved.

For unresolved sources, the *Strehl ratio* is defined as the ratio of observed peak flux over true peak flux. In case of Gaussian random phase errors, the Strehl ratio R is related to the RMS phase error σ_ϕ by [42]

$$R = \exp\left(-\frac{\sigma_\phi^2}{2}\right). \tag{6.6}$$

A larger peak flux is equivalent to a smaller RMS phase error. This statement is more generally true, because all phase errors cause scattering of source power into sidelobes.

A change in the apparent source shape due to ionospheric phase errors leads to an increase in residual sidelobes after deconvolution. Deconvolution subtracts a time-averaged source image model from the visibility data at all time stamps. In the presence of time-variable phase errors, the mean source model deviates from the apparent, instantaneous sky emission and subtraction is incomplete. Residual sidelobes increase the RMS background noise level and, due to its non-Gaussian character, introduce structure into the image that mimics real sky emission. In LF observations, due to the scaling relation of the dirty beam with frequency (width $\propto \nu^{-1}$), residual sidelobes around bright sources can be visible at significant distances from the source.

6.2.3 Ionospheric Phase Calibration

Lonsdale [9] discussed four different regimes for (instantaneous) ionospheric phase calibration, depending on the different linear spatial scales involved. These scales are the

array size A, the scale size S of ionospheric phase fluctuations and the projected size V of the field-of-view (FoV) at a typical ionospheric height. We use the term *compact* array when $A \ll S$ and *extended* array when $A \gtrsim S$. Note that these definitions change with ionospheric conditions, so there is no fixed linear scale that defines the difference between compact and extended. A schematic overview of the different regimes is given in Figure 6.1.

The combination $AV/\mathbf{S}^2$ is a measure of the complexity of ionospheric phase calibration. Both S and V depend on the observing frequency ν. For a power-law spectral density of free electron density fluctuations (see Section 6.2.1) S scales with ν, and for a fixed circular antenna aperture V scales with ν^{-1}. Therefore, $AV/\mathbf{S}^2$ scales with ν^{-3}, signalling a rapid increase in calibration problems towards low frequencies.

Under *isoplanatic* conditions ($V \ll S$), the ionospheric phase error per antenna does not vary with viewing direction within the FoV, for both compact and large arrays (Lonsdale regimes 1 and 2, respectively). Phase-only self-calibration on short enough timescales is sufficient to remove the ionospheric phase errors from the visibilities.

Under *anisoplanatic* conditions ($V \gtrsim S$), the ionospheric phase error varies over the FoV of each antenna. A single phase correction per antenna is no longer sufficient. Self-calibration may still converge, but the resulting phase correction per antenna is a flux-weighted average of ionospheric phases across the FoV (see Section 6.3.1). Accurate self-calibration and imaging of individual very bright and relatively compact sources is therefore possible, even with extended arrays (see [61] for an example). For a compact array (Lonsdale regime 3), the FoV of different antennas effectively overlap at ionospheric height. The LoS of different antennas towards one source run close and parallel through the ionosphere. For an extended array (Lonsdale regime 4), the FoV of different antennas may partially overlap at ionospheric height, but not necessarily. Individual LoS from widespread antennas to one source may trace very different paths through the ionosphere

In regime 3, ionospheric phases behave as a spatial gradient over the array that varies with viewing direction. This causes the apparent position of sources to change with time and viewing direction, but no source deformation takes place. The 3-dimensional phase structure of the ionosphere can be effectively reduced to a 2-dimensional phase screen, by integrating the free electron density along the LoS (Equation 6.4). Radio waves that pass the virtual screen experience an instantaneous ionospheric phase rotation depending on the *pierce point* position (where the LoS pierces the phase screen). When assuming a fixed number of required ionospheric parameters per unit area of phase screen, calibration of a compact array requires a minimal number of parameters because each antenna illuminates the same part of the phase screen.

In regime 4, the dependence of ionospheric phase on antenna position and viewing direction is more complex. This causes source position shifts and source shape deformations that both vary with time and viewing direction. A 2-dimensional phase screen model may still be used, but only when the dominant phase fluctuations originate from a restricted height range $\Delta h \ll S$ in the ionosphere. The concept of a thin layer at a given height is attractive, because it reduces the complexity of the calibration problem drastically. When using an airmass function to incorporate a zenith angle dependence,
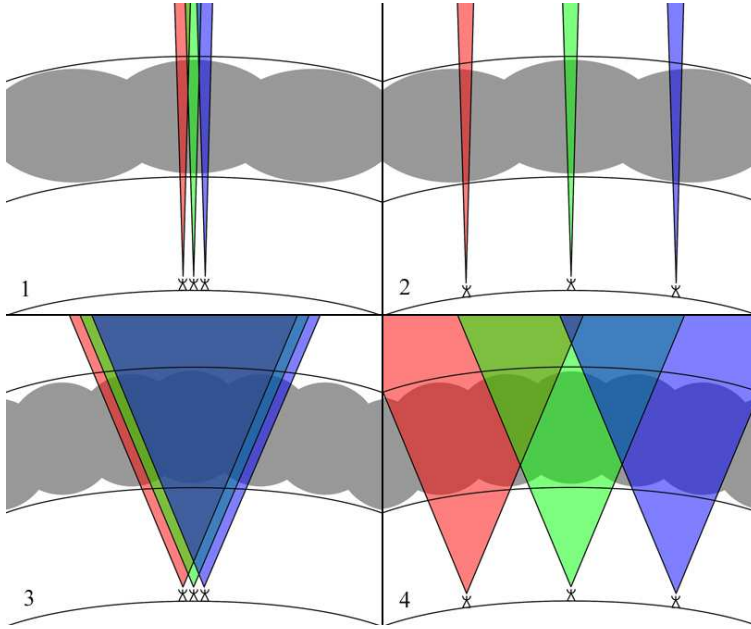
Figure 6.1: *Schematic overview of the different calibration regimes as discussed by Lonsdale [9]. For clarity, only two spatial dimensions and one calibration time interval are considered. In this overview, the array is represented by three antennas at ground level, looking through the ionospheric electron density structure (grey bubbles) with individual fields-of-view (red, green and blue areas). Due to the relatively narrow primary beam patterns in regimes 1 and 2 (top left and top right, respectively), each individual antenna 'sees' an approximately constant TEC across the FoV. The relatively wide primary beam patterns in regimes 3 and 4 (bottom left and bottom right, respectively) causes the antennas to 'see' TEC variations across the FoV. For the relatively compact array configurations in regimes 1 and 3, the TEC variation across the array for a single viewing direction within the FoV is approximately a gradient. For the relatively extended array configurations in regimes 2 and 4, the TEC variation across the array for a single viewing direction differs significantly from a gradient. The consequences for calibration of the array are discussed in the text.*

the spatial phase function is in effect reduced to 2 spatial dimensions. Generally, a phase screen in regime 4 requires a larger number of model parameters than in regime 3, because the phase screen area illuminated by the total array is larger.

It is currently unclear under which conditions a 2-dimensional phase screen model becomes too inaccurate to model the ionosphere in regime 4. For very long baselines or very severe ionospheric conditions, a full 3-dimensional ionospheric phase model may be required, where ionospheric phase corrections need to determined by ray-tracing. Such a model is likely to require many more parameters than can be extracted from radio observations alone. To first order, it may be sufficient to extend the phase screen model with some form of height-dependence. Examples of such extensions are the use of several phase screens at different heights [62] or introducing smoothly varying partial derivatives of TEC or phase as a function of zenith angle [63].

Calibration needs to determine corrections on sufficiently short time scales to track the ionospheric phase changes. The phase rate of change depends on the intrinsic time variability of the TEC along a given LoS and on the speed of the LoS from the array antennas through the ionosphere while tracking a cosmic source. The latter may range up to $\sim$ 100 km h^{-1} at 200 km height. The exact requirements on the time resolution of the calibration are yet to be determined. In principle, the time-variable ionospheric phase distortions needs to be sampled at least at the Nyquist frequency. However, during phase variations of large amplitude ($\gg$ 1 radian), 2π radian phase winding introduces periodicity on much shorter time scales. To succesfully unwrap phase winds, at least two corrections per 2π radian phase change are required.

6.2.4   Proposed and Existing Ionospheric Calibration Schemes

Schwab [64] and Subrahmanya [65] have proposed modifications to the self-calibration algorithm to support direction-dependent phase calibration. Both methods discuss the use of a spatial grid of interpolation nodes (additional free parameters) to characterize the spatial variability of the ionospheric phase error. Schwab suggests to use a different set of nodes per antenna, while Subrahmanya suggests to combine these sets by positioning them in a quasi-physical layer at fixed height above the Earth's surface (this to reduce the number of required nodes when the FoVs from different antennas overlap at ionospheric height). Neither of both proposed methods have been implemented.

Designed to operate in Lonsdale regime 3, field-based calibration by Cotton et al. [42] is the single existing implementation of a direction-dependent ionospheric phase calibration algorithm. Typically, for each time interval of 1–2 minutes of VLA 74 MHz data, the method measures and converts the apparent position shift of 5–10 detectable bright sources within the FoV into ionospheric phase gradients over the array. To predict phase gradients in arbitrary viewing directions for imaging of the full FoV, an independent phase screen per time interval is fitted to the measured phase gradients. The phase screen is described by a 5 term basis of Zernike polynomials (up to second order, excluding the constant zero order).

Field-based calibration has been used to calibrate 74 MHz VLA observations, mostly in B-configuration [41, e.g.] but also several in A-configuration [66, 67, e.g.]. Image

plane comparison of field-based calibration against self-calibration shows an overall increase of source peak fluxes (in some cases up to a factor of two) and reduction of residual sidelobes around bright sources, a clear indication of improved phase calibration over the FoV [55]. The improved overall calibration performance sometimes compromises the calibration towards the brightest source.

Zernike polynomials are often used to describe aberrations in optical systems, because lower order terms match well with several different types of wavefront distortions, and the functions are an orthogonal set on the circular domain of the telescope pupil. Using Zernike polynomials to describe an ionospheric phase screen may be less suitable, because they are not orthogonal on the discrete domain of pierce points, diverge when moving away from the field center and have no relation to ionospheric image abberations (except for first order, which can model a large scale TEC gradient). Non-orthogonality leads to interdependence between model parameters, while divergence is clearly non-physical and leads to undesirable extrapolation properties.

For extended LF arrays or more severe ionospheric conditions, the ionospheric phase behaviour over the array for a given viewing direction is no longer a simple gradient. Under these conditions, performance of field-based calibration degrades. For the 74 MHz VLA Low-frequency Sky Survey (VLSS; [41]), field-based calibration was unable to calibrate the VLA in B-configuration for about 10–20% of the observing time due to severe ionospheric conditions. Observing at 74 MHz with the $\sim$ 3 times larger VLA A-configuration leads to a relative increase in the failure rate of field-based calibration. This is to be expected, as the larger array size results in an increased probability for the observations to reside in Lonsdale regime 4.

The presence of higher order phase structure over the array in the direction of a calibrator requires an antenna-based phase calibration rather than a source position shift to measure ionospheric phases. The calibration methods proposed by Schwab and Subrahmanya (see above) do allow for higher order phase corrections over the array and could, in principle, handle more severe ionospheric conditions. An alternative approach is to use the *peeling* technique [26], which consist of sequential self-calibrations on individual bright sources in the FoV. This yields per source a set of time-variable antenna-based phase corrections and a source model. Because the peeling corrections are applicable to a limited set of viewing directions, they need to be interpolated in some intelligent way to arbitrary viewing directions while imaging the full FoV. Peeling is described in more detail in Section 6.3.3

Noordam [26] has proposed a 'generalized' self-calibration method for LOFAR [e.g. 68] that includes calibration of higher order ionospheric phase distortions. Similar to 'classical' self-calibration, instrumental and environmental (including ionospheric) parameters are estimated by calibration against a sky brightness model. Sky model and calibration parameters are iteratively updated to converge to some final result. Uniqueness of the calibration solution is controlled by putting restrictions on the time-, space- and frequency behaviour of the fitted parameters. The effects of the ionosphere are modeled in a Minimum Ionospheric Model (MIM; [63]), which is yet to be defined in detail. The philosophy of the MIM is to use a minimal number of physical assumptions and free parameters to accurately reproduce the observed effects of the ionosphere on the visi-

bilities for a wide-as-possible range of ionospheric conditions. The initial MIM is to be constrained using peeling corrections.

6.3   Method

SPAM, an abbreviation of 'Source Peeling and Atmospheric Modeling', is the implementation of a new ionospheric calibration method, combining several concepts from proposed and existing calibration methods. SPAM is designed to operate in Lonsdale regime 4 and can therefore also operate in regimes 1 to 3. It uses the calibration phases from peeling sources in the FoV to constrain an ionospheric phase screen model. The phase screen mimics a thin turbulent layer at a fixed height above the Earth's surface, in concordance with the observations of ionospheric small-scale structure (Section 6.2.1). The main motivation for this work was to test several aspects of ionospheric calibration on existing VLA and GMRT data sets on viability and qualitative performance, and thereby support the development of more advanced calibration algorithms for future instruments such as LOFAR.

Generally, the instantaneous ionosphere can only be sparsely sampled, due to the non-uniform sky distribution of a limited number of suitable calibrators and an array layout that is optimized for UV-coverage rather than ionospheric calibration. To minimize the error while interpolating to unsampled regions, an optimal choice of base functions for the description of the phase screen is of great importance. Based on the results in Chapter 5 [see also 51], we use the discrete Karhunen-Loève (KL) transform to determine an optimal set of base 'functions' to describe our phase screen. For a given pierce point layout and an assumed power-law slope for the spatial structure function of ionospheric phase fluctuations (see Section 6.2.1), the KL transform yields a set of base vectors with several important properties: (i) the vectors are orthogonal on the pierce point domain, (ii) truncation of the set (reduction of the model order) gives a minimal loss of information, (iii) interpolation to arbitrary pierce point locations obeys the phase structure function, and (iv) spatial phase variability scales with pierce point density, i.e., most phase screen structure is present in the vicinity of pierce points, while it converges to zero at infinite distance (more detail on this phase screen model is given in Section 6.3.4).

Because the required calibration time resolution is still an open issue, and the SPAM model does not incorporate any restrictions on temporal behaviour, independent phase screens are determined at the highest possible time resolution (which is the visibility integration time resolution).

SPAM calibration can be separated in a number of functional steps, each of which is discussed in detail in the sections to follow. The required input is a spectral-mode visibility data set that has flux calibration and bandpass calibration applied, and radio frequency interference (RFI) excised (see [69] or [41] for details). The SPAM recipe consists of the following steps:

1. Obtain and apply instrumental calibration corrections for phase (Section 6.3.1).

2. Obtain an initial model of the apparent sky, together with an initial ionospheric phase calibration (Section 6.3.2).

3. Subtract the sky model from the visibility data while applying the phase calibration. Peel apparently bright sources (Section 6.3.3).

4. Fit an ionospheric phase screen model to the peeling solutions (Section 6.3.4).

5. Apply the model phases on a facet-to-facet basis during re-imaging of the apparent sky (Section 6.3.5).

Steps 3 to 5 define the SPAM calibration cycle, as the image produced in step 5 can serve as an improved model of the apparent sky in step 3.

The scope of applications for SPAM is limited by a number of assumptions that were made to simplify the current implementation:

• The ionospheric inhomogeneities that cause significant phase distortions are located in a single, relatively narrow height range.

• There exists a finitely small angular patch size, which can be much smaller than the FoV of an individual antenna, over which the ionospheric phase contribution is effectively constant. Moving from one patch to neighbouring patches results in small phase transitions ($\ll 1$ radian).

• There exists a finitely small time range, larger than the integration time interval of an observation, over which the apparent ionospheric phase change for any of the array antennas along any line-of-sight is much smaller than a radian.

• The bandwidth of the observations is small enough to be effectively monochromatic, so that the ionospheric dispersion of waves within the frequency band is negligible.

• Within the given limitations on bandwidth and integration time, the array is sensitive enough to detect at least a few ($\gtrsim 5$) sources within the target FoV that may serve as phase calibrators.

• The ionospheric conditions during the observing run are such that self-calibration is able to produce a good enough initial calibration and sky model to allow for peeling of multiple sources. This might not work under very bad ionospheric conditions, but for the applications presented in this article it proved to be sufficient.

• After each calibration cycle (steps 3 to 5), the calibration and sky model are equally or more accurate than the previous. This implies convergence to a best achievable image.

• The instrumental amplitude and phase contributions to the visibilities, including the antenna power patterns projected onto the sky towards the target source, are constant over the duration of the observing run.

SPAM does not attempt to model the effects of ionospheric Faraday rotation on polarization products, and is therefore only applicable to intensity measurements (stokes I).

In our implementation we have focussed on functionality rather than processing speed. In its current form, SPAM is capable of processing quite large offline data sets, but is not suitable for real-time processing as is required for LOFAR calibration. SPAM relies heavily on functionality available in NRAO's Astronomical Image Processing System (AIPS; [e.g. 70]). It consists of a collection of Python scripts that accesses AIPS tasks, files and tables using the ParselTongue interface [71]. Two main reasons to use AIPS are its familiarity and proven robustness while serving a large group of users over a 30 year lifetime, and the quite natural way by which the ionospheric calibration method is combined with polyhedron imaging [72, 73]. SPAM uses a number of 3rd party Python libraries, like scipy, numpy and matplotlib for math and matrix operations and plotting. For non-linear least squares fitting of ionospheric phase models, we have adopted a Levenberg-Marquardt solver (LM; [e.g. 74]) based on IDL's MPFIT package [75].

6.3.1   Instrumental Phase Calibration

Each antenna in the array adds an instrumental phase offset to the recorded signal before correlation. At low frequencies, changes in the instrumental signal path length (e.g., due to temperature induced cable length differences) are very small compared to the wavelength, therefore instrumental phase offsets are generally stable over long time periods (hours to days). SPAM requires removal of the instrumental phase offsets from the visibilities prior to ionospheric calibration.

Instead of directly measuring the sky intensity $I(l, m)$ as a function of viewing direction cosines (l, m), an interferometer measures an approximate Fourier transform of the sky intensity. For a baseline consisting of antennas i and j, the perfect response to all visible sky emission for a single time instance and frequency is given by the measurement equation (ME) for visibilities [4, e.g.]:

$$V_{ij} = \int \int I(l, m) e^{-2\pi J \left[u_{ij}l + v_{ij}m + w_{ij}(n-1)\right]} \frac{\mathrm{d}l \, \mathrm{d}m}{n}, \tag{6.7}$$

where J indicates the imaginary part of a complex number, $n = \sqrt{1 - l^2 - m^2}$, u_{ij} and v_{ij} are baseline coordinates in the UV plane (expressed in wavelengths) parallel to l and m, respectively, and w_{ij} is the perpendicular baseline coordinate along the LoS towards the chosen celestial *phase tracking center* at $(l, m) = (0, 0)$. In practise, these measurements are modified with predominantly antenna-based complex gain factors a_i that may vary with time, frequency, antenna position and viewing direction. This modifies the ME into

$$\hat{V}_{ij} = \int \int a_i(l, m) \, a_j^\dagger(l, m)$$
$$I(l, m) e^{-2\pi J \left[u_{ij}l + v_{ij}m + w_{ij}(n-1)\right]} \frac{\mathrm{d}l \, \mathrm{d}m}{n}. \tag{6.8}$$

Determination of the gain factors is generally referred to as *calibration*. When known,

only gain factors that do not depend on viewing direction can be removed from the visibility data prior to image reconstruction by applying the calibration:

$$V_{ij} = (a_i a_j^\dagger)^{-1} \hat{V}_{ij} \tag{6.9}$$

This operation is generally not possible for gain factors that do depend on viewing direction, because these gain factors cannot be moved in front of the integral in Equation 6.8. One may still choose to apply gain corrections for a single viewing direction (e.g. to image a particular source), but the accuracy of imaging and deconvolution of other visible sources will degrade when moving away from the selected viewing direction. A solution for wide-field imaging and deconvolving in the presence of direction-dependent gain factors is discussed in Section 6.3.5.

The standard approach for instrumental phase calibration at higher frequencies is to repeatedly observe a bright (mostly unresolved) source during an observing run. Antenna-based gain phase corrections $g_i \approx a_i^{-1}$ are estimated by minimizing the weighted difference sum S between observed visibilities $\hat{V}_{ij}$ and source model visibilities $V_{ij}^{\mathrm{model}} \approx V_{ij}$ [e.g. 4]; implemented in AIPS task CALIB):

$$S = \sum_i \sum_{j>i} W_{ij} \| V_{ij}^{\mathrm{model}} - g_i g_j^\dagger \hat{V}_{ij} \|^p, \tag{6.10}$$

with W_{ij} the visibility weight (reciproke of the uncertainty in the visibility measurement), $g_i = e^{i\phi_i^{\mathrm{cal}}}$ and p the power of the norm (typically 1 or 2). The source model visibilities V_{ij}^{model} are calculated using Equation 6.7 with $I(l, m) = I^{\mathrm{model}}(l, m)$. The phase corrections ϕ_i^{cal} consist of an instrumental and an atmospheric part. The corrections are interpolated in time and applied to the target field visibilities, under the assumptions that the instrumental and atmospheric phase offsets vary slowly in time, and that the atmospheric phase offsets in the direction of the target are equal to those in the direction of the calibrator.

At low frequencies, there are two complicating factors for the standard approach: (i) the FoV around the calibrator source is large and includes many other sources, and (ii) the ionospheric phase offset per antenna changes significantly with time and viewing direction. The former can be overcome by choosing a very bright calibrator source with a flux that dominates over the combined flux of all other visible sources on all baselines (the gains g_i are a). For the VLSS [41], the 17,000 Jy of Cygnus A was more than sufficient to dominate over the total apparent flux of $400 - 500$ Jy in a typical VLSS field. The latter requires filtering of the phase corrections to extract only the instrumental part, which is then applied to the target field visibilities.

For SPAM, we have adopted an instrumental phase calibration method that is very similar to the procedure used for field-based calibration [42]. Antenna-based phase corrections are obtained on the highest possible time resolution by calibration on a very bright source k using the robust L1 norm (Equation 6.10 with $p = 1$; [76]). A phase correction $\phi_{ikn}^{\mathrm{cal}}$ for antenna i at time interval n consist of several contributions:

$$\phi_{ikn}^{\mathrm{cal}} = \phi_i^{\mathrm{instr}} + \phi_{ikn}^{\mathrm{ion}} - \phi_{rkn} - \phi_{ikn}^{\mathrm{ambig}}, \tag{6.11}$$

where the instrumental and ionospheric phase corrections, ϕ_i^{instr} and $\phi_{ikn}^{\mathrm{ion}}$ respectively, are assumed to be constant resp. vary with time and antenna position over the observing run. The other right-hand terms are the phase offset $\phi_{rkn} = \phi_r^{\mathrm{instr}} + \phi_{rkn}^{\mathrm{ion}}$ of an arbitrarily chosen reference antenna $r \in \{i\}$, and the phase ambiguity term $\phi_{ikn}^{\mathrm{ambig}} = 2\pi N_{ikn}$ with integer N_{ikn} that maps $\phi_{ikn}^{\mathrm{cal}}$ into the $[0, 2\pi)$ domain.

The antenna-based phase corrections are split into instrumental and ionospheric parts on the basis of their temporal and spatial behaviour. The phase corrections are filtered by iterative estimation of invariant instrumental phases (together with the phase ambiguities) and time- and space-variant ionospheric phases. The instrumental phases are estimated by robust averaging ($+3\,\sigma$ rejection) over all time intervals n:

$$\tilde{\phi}_i^{\mathrm{instr}} = \left\langle \left(\phi_{ikn}^{\mathrm{cal}} - \tilde{\phi}_i^{\mathrm{ion}}\right) \bmod 2\pi \right\rangle_n. \tag{6.12}$$

The phase ambiguity estimates follow from

$$\tilde{\phi}_{ikn}^{\mathrm{ambig}} = 2\pi \operatorname{round}\left(\left[\tilde{\phi}_i^{\mathrm{instr}} + \tilde{\phi}_i^{\mathrm{ion}} - \phi_{ikn}^{\mathrm{cal}}\right]/2\pi\right), \tag{6.13}$$

where the round() operator rounds a number to the nearest integer value. The instrumental phase offset of the reference antenna is arbitrarily set to zero. The ionospheric phases are constrained by fitting a time-varying spatial gradient $\vec{G}_{kn}$ to the phases over the array. The gradient fit consists of an initial estimate directly from the calibration phase corrections, followed by a refined fit by using the LM solver to minimize

$$\chi_{kn}^2 \;=\; \sum_i \left[\left(\phi_{ikn}^{\mathrm{cal}} - \tilde{\phi}_i^{\mathrm{instr}} + \tilde{\phi}_{ikn}^{\mathrm{ambig}}\right) - \underbrace{\vec{G}_{kn} \cdot (\vec{x}_i - \vec{x}_r)}_{\tilde{\phi}_{ikn}^{\mathrm{ion}}} \right]^2, \tag{6.14}$$

where $\vec{x}_i$ is the position of antenna i. The ionospheric phase offset of the reference antenna is arbitrarily set to zero, which makes it a pivot point over which the phase gradient rotates. Higher order ionospheric effects are assumed to average to zero in Equation 6.12.

6.3.2 Initial Phase Calibration and Initial Sky Model

The instrumental phase calibration method described in Section 6.3.1 assumes that the time-averaged ionospheric phase gradient over the array in the direction of the bright phase calibrator is zero. Any non-zero average is absorbed into the instrumental phase estimates, causing a position shift of the whole target field and thereby invalidating the astrometry. Before entering the calibration cycle (Sections 6.3.3–6.3.5), SPAM requires restoration of the astrometry and determination of an initial sky model and initial ionospheric calibration.

To restore the astrometry, the instrumentally corrected target field data from Section 6.3.1 is phase calibrated against an apparent sky model (AIPS task CALIB). The

default is a point source model, using NVSS catalog positions [77, 78], power-law interpolated fluxes from NVSS and WENSS/WISH catalogs [79] and a given primary beam model. To preserve the instrumental phase calibration as obtained in Section 6.3.1 during further processing, time-variable phase corrections resulting from calibration steps in this and the following sections are stored in a table (AIPS SN table) rather than applied directly to the visibility data. The sky model calibration is followed by wide-field imaging (AIPS task IMAGR) and several rounds of phase-only self-calibration (CALIB and IMAGR) at the highest possible time resolution, yielding the initial sky model and initial phase calibration.

For wide-field imaging with non-coplanar arrays, the standard imaging assumptions that the relevant sky area is approximately flat and the third baseline coordinate (*w*-term in Equation 6.7) is constant across the FoV are no longer valid. To overcome this, SPAM uses the polyhedron method [72, 73] that divides the large FoV into a hexagonal grid of small, partially overlapping *facets* that individually do satisfy the assumptions above (AIPS task SETFC). Additional facets are centered on relatively bright sources inside and outside the primary beam area to reduce image artefacts due to pixellation [80–84].

The Cotton-Schwab algorithm [64, 85, 86] is a variant of CLEAN deconvolution [14, 87] that allows for simultaneous deconvolution of multiple facets, using a different dirty beam for each facet. *Boxes* are used to restrict CLEANing to real sky emission, making sure that sources are deconvolved in the nearest facet only (CLEAN model components are stored in facet-based AIPS CC tables). After deconvolution, the CLEAN model is restored to the relevant residual facets (AIPS task CCRES) using a CLEAN beam, and the facets are combined to form a single image of the full FoV (AIPS task FLATN).

6.3.3   Peeling

To construct a model of ionospheric phase rotations in arbitrary viewing directions within the FoV, SPAM requires measurements in as many directions as possible. When no external sources of ionospheric information are available, the target field visibilities themselves need to be utilized. (Self-)calibration on individual sources can supply the required information, even in the presence of higher order phase structure over the array. After instrumental phase offsets are removed, phase calibration corrections are an relative measure of ionospheric phase:

$$\phi_{ikn}^{\text{cal}} = \phi_{ikn}^{\text{ion}} - \phi_{rkn}^{\text{ion}} - \phi_{ikn}^{\text{ambig}}, \tag{6.15}$$

where we used Equation 6.11 with $\phi_i^{\text{instr}} = \phi_r^{\text{instr}} = 0$.

SPAM uses the peeling technique [26] to obtain phase corrections in different viewing directions. Peeling consists of self-calibration on individual sources, yielding per source a set of time-variable antenna-based phase corrections and a source model. After self-calibration of a source, the source model is subtracted from the visibility data set while temporarily applying the phase corrections (AIPS tasks SPLIT, UVSUB and CLINV/SPLIT).

For peeling to converge, a source needs to be the dominant contributor of flux to the visibilities on all baselines. Especially at low frequencies, the presence of many other

sources in the large FoV may add considerable noise to the peeling phase corrections. To suppress this effect, the following steps are performed: (i) The best available model of the apparent sky is subtracted from the visibility data while temporarily applying the associated phase calibration(s). The initial best available model and associated phase calibration is the self-calibration output of Section 6.3.2. Individual source models are added back before peeling. (ii) Sources are peeled in decreasing flux order to suppress the effect of brighter sources on the peeling of fainter sources. (iii) Calibration only uses visibilities with projected baseline lengths longer than a certain threshold. This excludes the high 'noise' in the visibilities near zero-length baselines from the coherent flux contribution of imperfectly subtracted sources.

The radio sky can be approximated by a discrete number of isolated, invariant sources of finite angular extend. Visibilities in the ME (Equation 6.7) for a single integration time n can therefore be split into a linear combination of contributions from individual sources k:

$$V_{ijn} = \sum_k V_{ijkn} = \sum_k \int \int I_k(l,m)$$
$$e^{-2\pi J\left[u_{ijn}l + v_{ijn}m + w_{ijn}(n-1)\right]} \frac{\mathrm{d}l\,\mathrm{d}m}{n}. \tag{6.16}$$

The subtraction of all but the peeling source k' from the measured visibilities in step (i) above can be described as

$$\hat{V}_{ijk'n} \approx \hat{V}_{ijn} - \sum_{k \neq k'} (g_{ikn}\,g_{jkn}^{\dagger})^{-1} V_{ijkn}^{\text{model}}, \tag{6.17}$$

with $g_{ikn} = g_i(l_k, m_k, t_n) = e^{i\phi_{ikn}^{\text{cal}}}$ the best available calibration in the viewing direction of source k, and V_{ijkn}^{model} the visibilities that are derived from the best available model I_{ijk}^{model} of source k. The peeling itself consists of iterative calibration and imaging steps of the peeling source k'. The calibration (Equation 6.10 with $p = 1$) updates the antenna gain corrections g_{ikn} by minimizing

$$S_n = \sum_i \sum_{j>i} w_{ijn} \|V_{ijk'n}^{\text{model}} - g_{in}\,g_{jn}^{\dagger}\hat{V}_{ijk'n}\|, \tag{6.18}$$

while the imaging step updates $I_{ijk'}^{\text{model}}$ and therefore $V_{ijk'n}^{\text{model}}$.

In practise, due to incompleteness of the sky model and inaccuracies in the phase calibration, there will always remain some contaminating source flux in the visibilities while peeling. Complemented with system noise, sky noise, residual RFI and other possible sources of noise, the noise in the visibilities propagates into the phase corrections from the peeling process.

Absolute astrometry is not conserved during peeling, because self-calibration allows antenna-based phase corrections to vary without constraint. In subsequent peeling cycles, small non-zero phase gradients in the phase residuals after calibration can cause the

source model to wander away from its true position. In SPAM, astrometry errors are minimized by re-centering the source model to its true (catalog) position before calibration in each self-calibration loop. By default, SPAM re-centers the peak of the model flux to the nearest bright point source position in the NVSS catalog [77, 78]. It is recommended to visually check the final peeling source images for possible mismatches with the catalog (e.g., in case of double sources or sources with a spatially varying spectral index).

While peeling, SPAM attempts to calibrate sources on the highest possible time resolution, which is the visibility time grid. The noise in the resulting phase corrections depends on the signal-to-noise ratio (SNR) of the source flux in the visibilities. To increase the number of peeling sources and limit the phase noise in case of insufficient SNR, SPAM is allowed to increase the calibration time interval beyond the visibility integration time up to an arbitrary limit. Through image plane analysis, SPAM estimates the required calibration time-interval per source:

$$n_t = \left(\frac{\sigma_L}{\alpha S_p}\right)^2 N_t,$$ (6.19)

where n_t is the required number of integration times in a calibration interval, N_t is the total number of integration times within the observation, α is the minimum required SNR per integration time (a tweakable parameter that sets the balance between the SNR and the time resolution of the peeling phase corrections), and S_p and σ_L are the measured source peak flux and local background noise level in the image. For a fixed upper limit on the calibration time interval, an increase in α results in a decrease in the number of peeling sources. For $n_t < 1$, phase corrections are determined on the visibility time grid. For $n_t > 1$, a spline is used to resample the phase corrections per antenna in time onto the visibility time grid.

Apart from SNR issues, the number of sources that can be peeled is fundamentally limited by the available number of independent visibility measurements. When peeling N_s sources, self-calibration fits $N_s(N_a - 1)$ phase solutions per calibration time interval to the visibility data, where N_a is the number of antennas. For self-calibration to converge to an unique combination of phase solutions and source model, this number needs to be much smaller than the number of independent visibility measurements. The maximum of visibilities measurements that is available in one calibration time interval is given by $N_c \langle n_t \rangle N_a (N_a - 1)/2$, with N_c the number of frequency channels and $\langle n_t \rangle$ the average number of visibility integration times in a calibration interval. In the ideal case, when we assume that each visibility is an independent measurement, the determination of antenna-based phase corrections for all peeling sources is well constrained if

$$N_s \ll \frac{N_a N_c \langle n_t \rangle}{2}.$$ (6.20)

The applications presented in this article do satisfy this minimal condition (see Section 6.4).

Equation 6.20 is equivalent to stating that the number of degrees-of-freedom (DoF; the difference between the number of independent measurements and the number of

model parameters) should remain a large positive number. Correlation between visibilities over frequency and time may reduce the number of independent measurements drastically, thereby also reducing the number of DoFs. The exact number of DoFs for any data set is hard to quantify. When this number becomes too low, the data is 'over-fitted' [e.g. 88], which could result in an artificial reduction of both the image background noise level and source flux that is not represented in the self-calibration model [12]. Although we have found no evidence of this effect occuring in the applications presented in this article, the SPAM user should be cautious not to peel too many sources. In case of a high number of available peeling sources, one can choose a subset with a sufficiently dense spatial distribution over the FoV (e.g., one source per isoplanatic patch; see Section 6.3.5).

6.3.4 Ionospheric Phase Screen Model

The phase corrections that are obtained by peeling several bright sources in the FoV (Section 6.3.3) are only valid for ionospheric calibration in a limited patch of sky around each source. To correct for ionospheric phase errors over the full FoV during wide-field imaging and deconvolution, SPAM requires a model that predicts the phase correction per antenna in arbitrary viewing directions.

SPAM constructs a quasi-physical phase screen model that attempts to accurately reproduce and interpolate the measured ionospheric phase rotations (or more accurately: the peeling phase corrections). The phase screen is determined independently for each visibility time stamp, therefore we drop the n-subscript in the description below. Figure 6.2 is a schematic overview of the geometry of ionospheric phase modeling in SPAM. The ionosphere is represented by a curved phase screen at a fixed height h above the Earth's surface, compliant to the WGS84 standard (NIMA [89]). The total phase rotation experienced by a ray of radio emission traveling along a LoS through the ionosphere is represented by an instantaneous phase rotation $\phi^{\mathrm{ion}}(\vec{p}, \zeta)$ on passage through the phase screen that is a function of pierce point position $\vec{p}$ and zenith angle ζ. For a thin layer ($\Delta h \ll S$; see Section 6.2.3), the dependence of ϕ^{ion} on ζ can be represented by a simple airmass function, so that

$$\phi^{\mathrm{ion}}(\vec{p}, \zeta) = \frac{\phi^{\mathrm{ion}}(\vec{p})}{\cos(\zeta)}. \tag{6.21}$$

SPAM uses an angular local longitude/latitude coordinate system to specify $\vec{p}$, relative to the central pierce point from array center to field center. For the applications presented in this article, the angular distances between pierce points over the relevant ionospheric domain are all < 5 degrees, which effectively makes the pierce point vector $\vec{p}$ a 2-dimensional cartesian vector.

The 2-dimensional phase screen $\phi^{\mathrm{ion}}(\vec{p})$ is defined on a set of KL base vectors, generated from the instantaneous pierce point configuration $\{\vec{p}_{ik}\}$ and an assumed power-law shape for the phase structure function (Section 6.2.1). The KL base vector generation and interpolation described in detail in Section 5.3.5. The phase screen model requires one free parameter per KL base vector. The initial complete set of KL base vectors is
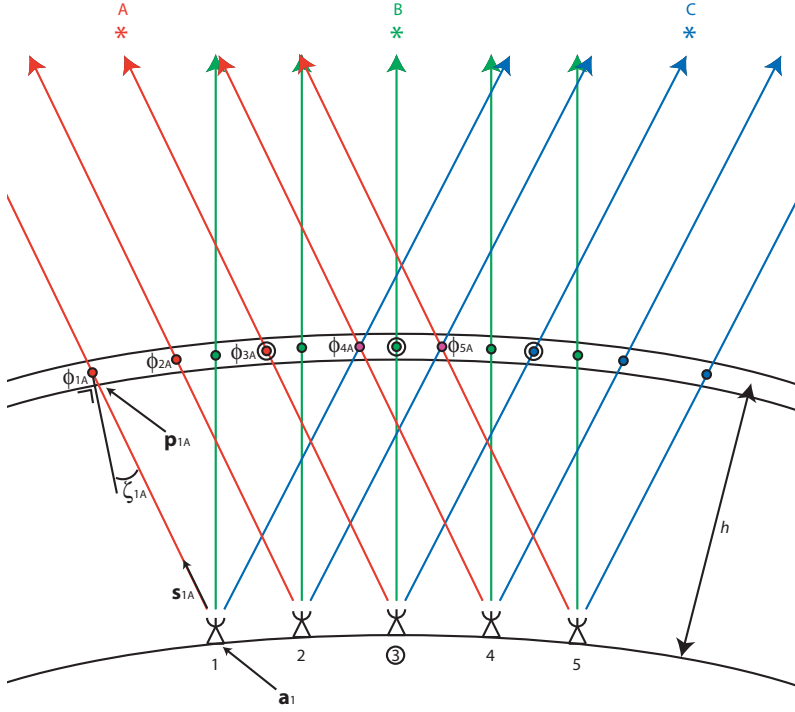
Figure 6.2: *Schematic overview of the SPAM thin ionospheric phase screen model geometry. For clarity, only two spatial dimensions and one calibration time interval are considered. In this overview, five ground-based array antennas (labelled 1 to 5) observe three calibrator sources (colored red/green/blue and labelled A to C) within the FoV. The (colored) LoSs from the array towards the sources run parallel for each source and pierce the phase screen at fixed height h (colored circles). The LoS from antenna i at Earth location $\vec{a}_i$ towards a peeling source k at local sky position $\hat{s}_{ik}$ intersects the phase screen at a single pierce point $\vec{p}_{ik}$ under a zenith angle ζ_{ik}. For a single LoS from antenna 1 towards source A, we have indicated how the pierce point position $\vec{p}_{ik} = \mathbf{p}_{1A}$ and zenith angle $\zeta_{ik} = \zeta_{1A}$ relate to the antenna position $\vec{a}_i = \mathbf{a}_1$ and the local sky position $\hat{s}_{ik} = \mathbf{s}_{1A}$ of the source. For some LoSs the pierce points may overlap (or nearly overlap), as is the case for 1C & 4A and 2C & 5A in our example. The total (integrated) phase rotation along any LoS through the ionosphere is modeled by an instantaneous phase rotation ϕ_{ik}^{ion} at the phase screen height. For example, radio waves traveling along LoSs from source A towards antennas 1 to 5 experience an instantaneous phase rotation $\phi_{ik}^{\mathrm{ion}} = \phi_{1A}$ to ϕ_{5A}, respectively, while passing the screen at their related pierce points $\vec{p}_{ik} = \mathbf{p}_{1A}$ to $\mathbf{p}_{5A}$, respectively. Peeling the three calibrator sources yields measurements of the ionospheric phases ϕ_{ik}^{ion}, relative to a common reference antenna (in this example antenna 3; encircled).*
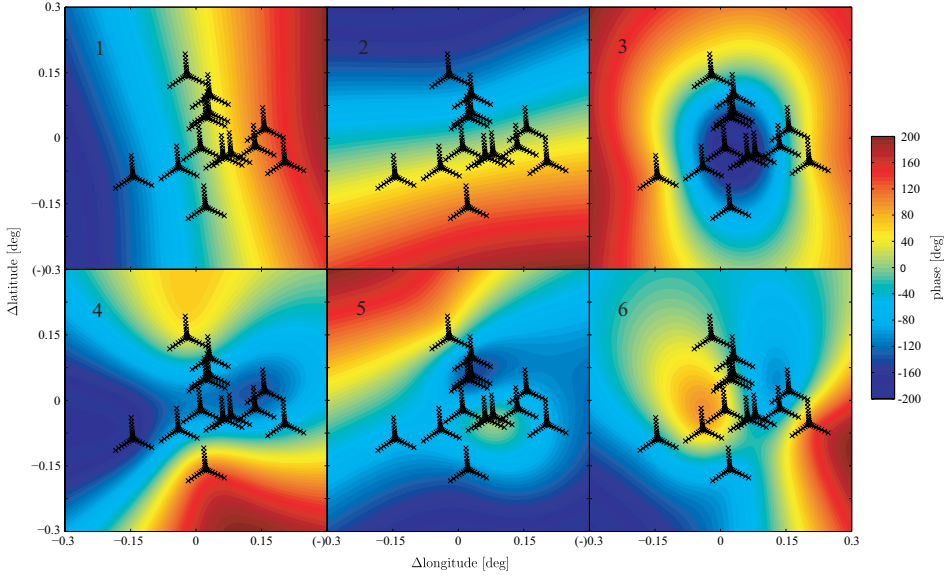
Figure 6.3: *Plots of the interpolations of the first six KL base vectors, derived for an artificial but realistic configuration of ionospheric pierce points. In this example, the pierce points (black crosses) are calculated for a single time instance during a 74 MHz VLA-B observation with 13 available calibrator sources in the ~ 10 degree FoV, adopting a phase screen height h = 200 km and a structure function power-law slope γ = 5/3. The horizontal and vertical axes represent angular distances in East-West and North-South directions, respectively, as seen from the center of the Earth, relative to the phase screen's pierce point along the line-of-sight from array center to pointing center, with East- and Northward offsets being positive. At this height, a 0.1 degree angular offset represents a physical horizontal offset of ~ 11.5 km. The direction-dependent phase for each interpolated KL base vector is color-coded and scaled to an arbitrary amplitude range.*

arbitrarily reduced in order by selecting a subset based on statistical relevance (principle component analysis). This reduces the effect of noise in the peeling solutions on the model accuracy and simultaneously limits the number of model parameters. However, the subset should still be large enough to accurately reproduce the peeling phase corrections. Per visibility time stamp, the KL base vectors are stored for later use during imaging (for this purpose, we mis-use the AIPS OB table). As an example, the first six interpolated KL base vectors for a single configuration of ionospheric pierce points are plotted in Figure 6.3.

The peeling phase corrections ϕ_{ik}^{cal} are interpreted to be relative measurements of the absolute ionospheric phase screen model $\phi^{\mathrm{ion}}(\vec{p}, \zeta)$ which may be determined up to a constant. The model parameters are determined by minimizing the differences between the observed and the model phases using the LM non-linear least-squares solver, for

which a χ^2 sum needs to be defined. From Equation 6.15 it follows that

$$\phi_{ik}^{\text{cal}} = \phi^{\text{ion}}(\vec{p}_{ik}, \zeta_{ik}) - \phi^{\text{ion}}(\vec{p}_{rk}, \zeta_{rk}) - \phi_{ik}^{\text{ambig}}. \tag{6.22}$$

Consequently, the phase correction in the direction of source k for a baseline consisting of antennas i and j is

$$\begin{aligned}
\phi_{ik}^{\text{cal}} - \phi_{jk}^{\text{cal}} &= [\phi^{\text{ion}}(\vec{p}_{ik}, \zeta_{ik}) - \phi^{\text{ion}}(\vec{p}_{jk}, \zeta_{jk})] - \\
&\qquad [\phi_{ik}^{\text{ambig}} - \phi_{jk}^{\text{ambig}}].
\end{aligned} \tag{6.23}$$

The χ^2 sum is defined as:

$$\begin{aligned}
\chi^2 &= \sum_k \sum_i \sum_{j>i} \Big[\big([\phi_{ik}^{cal} - \phi_{jk}^{cal}] - \\
&\qquad [\phi^{\text{ion}}(\vec{p}_{ik}, \zeta_{ik}) - \phi^{\text{ion}}(\vec{p}_{jk}, \zeta_{jk})] \big) \bmod 2\pi \Big]^2.
\end{aligned} \tag{6.24}$$

This definition has several properties: (i) By remapping the χ^2 terms into the $[0, 2\pi)$ domain, the phase ambiguity terms do not have to be fitted explicitly, (ii) the χ^2 terms of all calibrator sources are weighted equally, so the model is not biased towards the brightest source (as is the case for self-calibration), and (iii) using χ^2 terms from all possible antenna pairs prevents a bias towards the reference antenna.

Using Equation 6.24, the LM solver yields a set of model parameters per visibility time stamp. These are stored for later use during imaging (AIPS NI table). The square root of the average of the χ^2 terms equals the average RMS phase residual between peeling and model phases. Time intervals that have a bad fit are identified and removed by means of an upper limit ($+2.5\,\sigma$ rejection) on the distribution of RMS phase residuals over time.

Convergence of the LM solver is troubled by 2π phase ambiguities, because these introduce local minima in χ^2 space. A good initial guess of the model parameters greatly helps to overcome this problem. To this purpose, SPAM estimates the global phase gradient over all the pierce points directly from the phase corrections ϕ_{ik}^{cal} and projects it onto the KL base vectors before invoking the LM solver.

Figure 6.4 shows an example of an ionospheric phase screen that was constructed as described above. The pierce point layout consists of multiple projections of the array onto the phase screen. The low density of calibrators causes a minimal overlap between array projections. Figure 6.5 shows a comparison between time-sequences of phase corrections from self-calibration, peeling and model fitting. Because the self-calibration corrections are a flux-weighted average for the full FoV, they are biased towards the brightest source. They look somewhat similar to the peeling solutions of the brightest source, but the latter contains additional fluctuations that vary on a relatively short timescale. The model phases appear similar to the peeling phases, but vary more smoothly. Their values fall somewhere in between the self-calibration phases and the peeling phases. The difference between the peeling phases and model phases are mainly caused by the constraints on the spatial variability of the phase screen model.
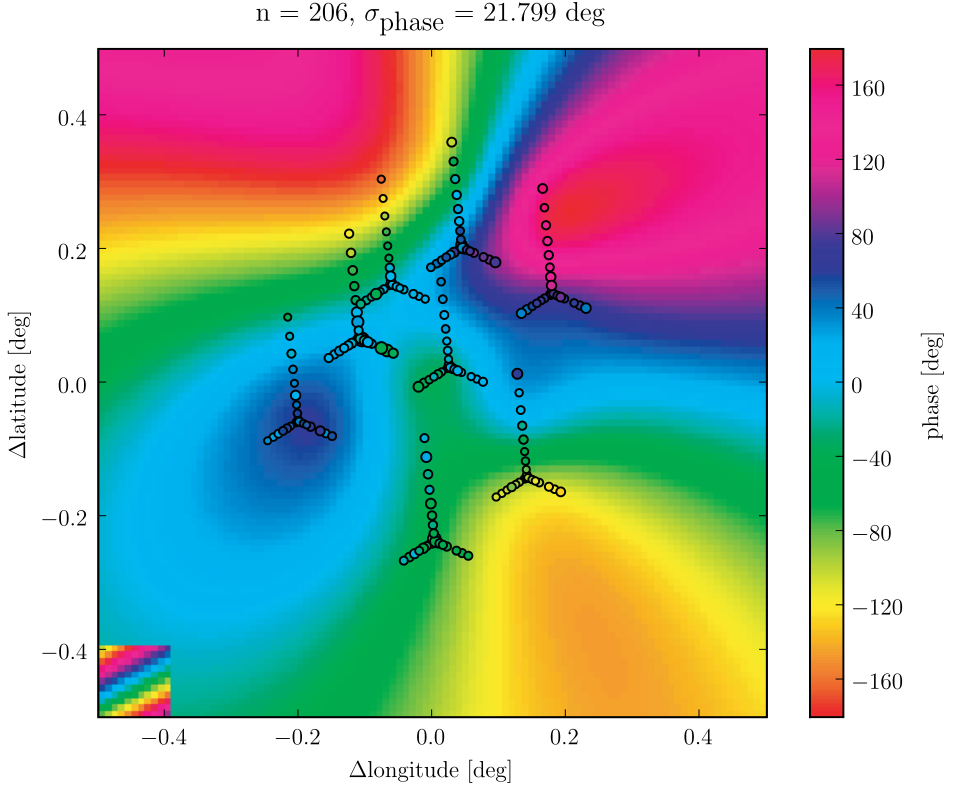
Figure 6.4: *Example of an ionospheric phase screen model fit. The color map represents an ionospheric phase screen at 200 km height that was fitted to the peeling phase solutions of 8 calibrator sources at time-interval n = 206 of 10 seconds during a VLSS observing run of the 74 MHz VLA in BnA-configuration (see Section 6.4, the J1300-208 data set). The plot layout is similar to Figure 6.3. The overall phase gradient (depicted in the bottom-left corner) was removed to make the higher order terms more clearly visible. The collection of pierce points from all array antennas to all peeling sources are depicted as small circles., The color in the circle represents the measured peeling phase (the reference antenna VLA N36 was set to match the phase screen value). The size of the circle scales with the magnitude of the estimated phase residual after model correction. The overall RMS phase residual σ_{phase} = 21.799 degrees (averaged over all pierce points) was one of the better fitting results during this particular observing run.*
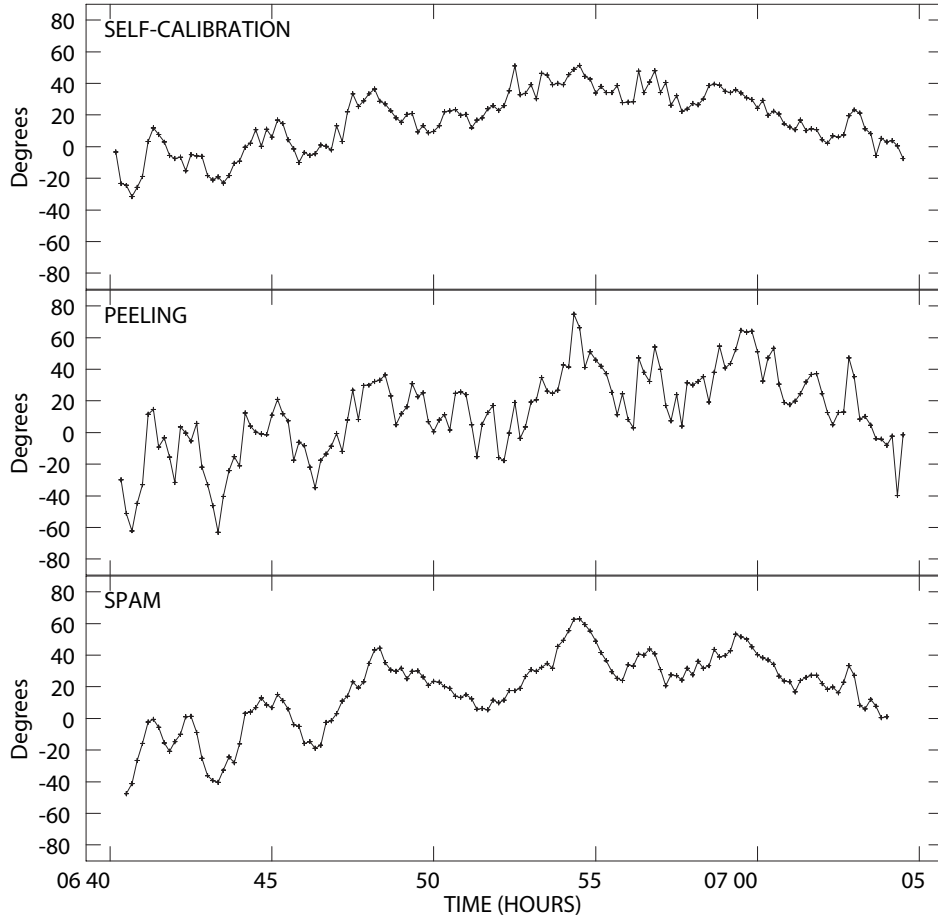
Figure 6.5: *Example of phase corrections from different steps in the ionospheric calibration process, resulting from processing a VLSS data set with SPAM (see Section 6.4, the real J0900+398 data set). The antenna under consideration is VLA E28, with W20 being the reference antenna (an 5.7 km east-west baseline). The plots represent 25 minutes of observing time, using a 10 second time resolution. Top: Antenna-based phase corrections resulting from self-calibration on the whole FoV. Middle: Phase corrections resulting from peeling the brightest (30 Jy) source. Bottom: Corrections resulting from ionospheric phase modeling in the direction of the (same) brightest source.*

6.3.5   Imaging

With an ionospheric phase screen model available for a given visibility data set, antenna-based phase corrections for any direction in the wide FoV can be calculated (Equation 6.22). Because each visibility consist of contributions from visible sources in different viewing directions, there is no simple operation that removes the ionospheric phase rotations from a visibility data set prior to imaging. Instead, SPAM requires an algorithm that calculates and applies the appropriate model phase corrections during imaging and deconvolving for different parts of the FoV.

SPAM works under the assumption that there exists a fixed angular *isoplanatic patch* size on the sky, with a projected size at ionospheric height smaller than the scale size of ionospheric phase fluctuations, over which variations in ionospheric phase rotation are negligible. Each isoplanatic patch requires at least one phase correction per antenna per visibility time interval. For the VLA at 74 MHz, the isoplanatic patch size is estimated to be 2–4 degrees [55].

The facet-based polyhedron method for wide-field imaging (see Section 6.3.2) allows for a relatively simple implementation of ionospheric phase correction [64]. By choosing a facet size smaller than the isoplanatic patch size, a set of model phase corrections calculated for the center of a facet are assumed to be accurate for the whole facet area. Ionospheric phase model corrections are calculated and stored (AIPS SN tables) for each facet center in the FoV prior to imaging and deconvolution. For the additional facets centered on bright sources (see Section 6.3.2), model phase corrections are optionally replaced by peeling phase corrections to allow for optimized calibration towards these sources.

The SPAM imaging and deconvolution procedure is similar to the procedure used for the field-based calibration method by Cotton et al. [42], which differs from the standard Cotton-Schwab algorithm by the temporary application of the facet-based phase corrections (AIPS tasks SPLIT and CLINV/SPLIT) to the visibility data for the duration of major CLEAN cycles on individual facets (AIPS tasks IMAGR and UVSUB). After deconvolution, facets are combined to form a single image of the full FoV (AIPS task FLATN). Because antenna-based phase corrections change very little between adjacent facets, the complete set of partly overlapping facet images combine into a continuous image of the FoV.

6.4   Applications

To demonstrate the capabilities of SPAM, we have defined three test cases based on observations with the VLA at 74 MHz [58]. In each test case, SPAM is used for ionospheric phase calibration and imaging of a VLSS visibility data set [41], following the steps described in Section 6.3. In the first test case, SPAM was applied to simulated data to validate basic functionality in a controlled environment. In the next two test cases, SPAM was applied to visibility data from real observations under varying ionospheric conditions. We compare SPAM performance against self-calibration (SC) and

field-based calibration (FBC) by analyzing the resulting images. The setup and results of these test cases are described in detail in the following sections.

6.4.1   Data Selection, Preparation and Processing

In this Section we describe how the visibility data sets for the three test cases were selected/constructed. Furthermore, we present details on how these data sets were processed by SPAM into calibrated images of the FoV.

Two VLSS observations, at pointing centers J0900+398 and J1300-208, respectively, have been picked from more than 500 available VLSS observations on the following criteria: (i) both fields contain a relatively large number of bright sources that can serve as calibrators, and (ii) the ionospheric conditions during the observations appear to be relatively good (J0900+398) and relatively bad (J1300-208). The presence of more than 5 bright sources of at least 5 Jy compensates for the relatively poor efficiency of the VLA 74 MHz receiving system [58]. The ionospheric conditions were derived from the apparent smearing of point sources in the images, due to residual phase errors after applying FBC. From experience, we adopted the qualification 'good' when the mean width of apparent point sources was at most $5''$ larger than the intrinsic $80''$ resolution, while for 'bad' conditions the mean point source width was larger by at least $15''$. In terms of Strehl ratio R (Equation 6.6), 'good' and 'bad' conditions correspond with $R > 0.996$ and $R < 0.966$, respectively. Additionally, candidate fields were visually inspected for evidence of residual phase errors by the presence or absence of image artefacts near bright sources, which lead to the final selection of the two fields mentioned above.

The difference in observed ionospheric conditions between the two real data sets may be the result of the difference in array size and elevation of the target field. From the VLA site at +34 degrees declination, the J0900+398 field was observed in B-configuration (up to 11 km baselines) at relatively high elevation, while the J1300-208 field was observed in BnA-configuration (up to 23 km baselines) at relatively low elevation. For the J1300-208 observation, the array observed through the ionosphere at larger separations and along longer path lengths than for the J0900+398 observation, which is expected to result in both larger and less coherent phase errors over the array.

Because both real data sets have been previously calibrated and imaged with FBC, the data sets were already partly reduced at the start of SPAM processing. Instrumental calibration was applied (including instrumental phase calibration, similar to Section 6.3.1), most RFI-contaminated data was flagged and the spectral resolution was reduced [see 41] for details), but no FBC has been applied yet. For the simulated data set, which is based on the real J0900+398 observations, the measured visibilities were replaced by noiseless model visibilities of an idealized sky, consisting of 91 bright point sources with peak fluxes (larger than 1 Jy) and positions as measured in the J0900+398 FBC image. For each point source, the corresponding model visibility phases were corrupted using the direction-dependent ionospheric phase model that was obtained with FBC to correct the real J0900+398 data.

FBC images of the two real data sets were available in the VLSS archive. For the simulated J0900+398 data set, an 'undisturbed' image was made before applying the iono-

spheric phase corruptions. All three VLSS data sets have been processed with SPAM, yielding both an SC image and an ionosphere-corrected SPAM image. Relevant details on the processing can be found in Table 6.1. For SC and SPAM imaging, we adopted most of the imaging-specific settings from FBC (like uniform weighting). Noticeable differences are the use of CLEAN boxes, a smaller pixel size and a different facet configuration.

By choosing a minimum SNR per time interval of 15 and a maximum peeling time interval of 4 minutes (see Equation 6.19), SPAM was able to peel ~ 10 sources in each of the real data sets. Lowering the SNR resulted in a much larger scatter in the peeling phases over time, or prevented peeling to converge at all. The peeling time upper limit was chosen to roughly match the spatial density of calibrator sources used in FBC. Determining phase corrections on a 4 minute time scale could result in undersampling the time evolution of ionospheric phase errors. Note that this only applies to the faintest of the calibrator sources. The limitations on spatial and temporal sampling of the ionosphere are dictated by the given sensitivity of the VLA.

Because of the high SNR, all 91 sources in the simulated J0900+398 data set qualified for peeling at the highest time resolution of 10 seconds. To mimic a more realistic scenario for further SPAM processing, the number of calibrators was arbitrarily limited to 10. Generally, for all data sets, the images of peeling sources showed larger peak fluxes and less background structure than their counterparts in the SC image, although the contrast became less apparent for weaker and extended (mostly doubles) peeling sources.

As stated in Section 6.3.3, the number of peeling sources is fundamentally limited by the requirement for a large positive number of degrees-of-freedom in the available visibility data. The minimal requirement is given in Equation 6.20. Typically, for the VLSS data sets, there were 25 active antennas, 12 frequency channels and 6 visibility intervals (of 10 seconds) in an average peeling interval of 1 minute. In our test cases, we typically peel 10 sources, which is much less than $25 \times 12 \times 6/2 = 900$, thereby satisfying the minimal requirement.

Due to the uncertainty in their optimal values, it is left to the SPAM user to specify the phase screen model order (the number of KL base vectors), the height h of the phase screen and the power-law exponent γ of the phase structure function. For the applications presented here, we used $h = 200$ km and $\gamma = 5/3$, which is compliant to the measured values given in Section 6.2.1 given the uncertainty in these values. For the simulated data set, we chose instead $h = 1000$ km to better match the corrupting FBC ionospheric phase model that is attached to the sky plane at infinite height. These values gave satisfactory results for the test applications presented here, but can be further optimized. The optimal model order was found to lie in the range of 15–20 terms, which is 1.5–2 times the number of available peeling sources. Increasing or decreasing the model order caused the model fit to be less accurate or more problematic in terms of convergence.

For both the simulated and real J0900+398 data sets, no improvement in background noise was observed by adding a second calibration cycle after the first. This indicates fast convergence of the SPAM calibration method for quiet ionospheric conditions, where the initial self-calibration is already close to the best achievable calibration of SPAM. For the real J1300-208 data set, adding up to third calibration cycle did improve over the previous

Field name	VLSS J0900+398 (simulated)	VLSS J0900+398 (real)	VLSS J1300-208 (real)
Pixel size[a]	18.9″	18.9″	11.1″
Number of facets	347	243	576
Facet separation	1.°18	1.°18	0.°62
SPAM calibration cycles[b]	1	1	3
Peeling sources	10[c]	11	9
KL model height	1000 km[d]	200 km	200 km
Fitted KL model terms	15	15	20[e]
Rejected time intervals	0 / 464	25 / 464	86 / 484
Model fit phase RMS	3.0 ± 0.8 degrees	21.3 ± 2.4 degrees	23.2 ± 3.2 degrees
Peeling corrections applied directly	no	yes	yes

[a] The pixel size for all field-based calibration images is 20″.

[b] Adding more cycles did not significantly improve the image quality.

[c] Arbitrarily limited to mimic a more realistic scenario.

[d] Increased to improve match with FBC phase screen.

[e] In this case, 15 terms proved to be insufficient.

Table 6.1: *Overview of processing parameters for the three data sets that are handled with SPAM as defined in the test cases.*

cycles.

6.4.2   Phase Calibration Accuracy

For the simulated J0900+398 data set, the absolute accuracy of ionospheric calibration can be determined by a direct comparison between the corrupting FBC phase screen and the correcting SPAM phase screen. To this purpose, phase corruptions and corrections were calculated from the models for a hexagonal grid of 342 viewing directions within the FoV. Per viewing direction, the RMS phase error was calculated by differencing of the phases from both models and averaging over all time stamps and baselines. The result is depicted in Figure 6.6.

For areas near the calibrators and in the center of the field in general, there is a relatively good match between the input and output model, with typical RMS phase errors $\lesssim 5$ degrees. The absence of calibrator sources south-west of the field center still results in relatively accurate predictions by the SPAM model. In the direction of peeling sources, the measured RMS phase error can be split into a contribution from inaccuracies in the peeling process and a contribution from imperfect model fitting. The latter is approximately 3 degrees (Table 6.1), therefore the RMS phase error introduced by peeling is $\lesssim 4$ degrees. Considering the model setup, the only possible source of error is contamination from other sources while peeling (which appears to happen despite the initial subtraction of the SC model).

Overall, the change in model base from the corrupting FBC model (5 Zernike polynomials) to the correcting SPAM model (15 KL vectors) has a constant accuracy over large parts of the FoV. Towards some parts of the edge of the field the phase errors are substantially larger, up to 20–25 degrees at worst. This agrees with the different asymptotic behaviour towards large radii of the Zernike model (diverge to infinity) and the KL model (converge to zero) in the absence of calibrators. The presence of calibrator sources near the edge (like the one on the North-East edge of the field) leads to a better local match between corrupting ionosphere and correcting model.

For the real observations, in the absence of external sources of information (e.g., GPS measurements), it is not possible to derive the absolute accuracy of ionospheric calibration from the observations themselves. Instead, the residual RMS phase error of the model fit to the peeling phases is used as an relative indicator for calibration accuracy over time. For both the real J0900+398 and J1300-208 data sets, the residual RMS phase error of ~ 22 degrees is much larger than for the simulated data. This already excludes rejected time stamps with exceptionally large RMS values. By inspecting model fits on individual time stamps, we found that there are often a few pierce point phases that deviate significantly more from the fitted model than most neighbouring points. These errors do not appear to be antenna-based instrumental errors, because peeling solutions for the same antenna towards other calibrator sources do not deviate in the same manner. Typically, these deviating points persist for a few time stamps before disappearing. The ionosphere may be responsible for these very small scale deviations. Another possibility is that the peeling solutions are (sometimes) noisy due to limitations in source SNR.
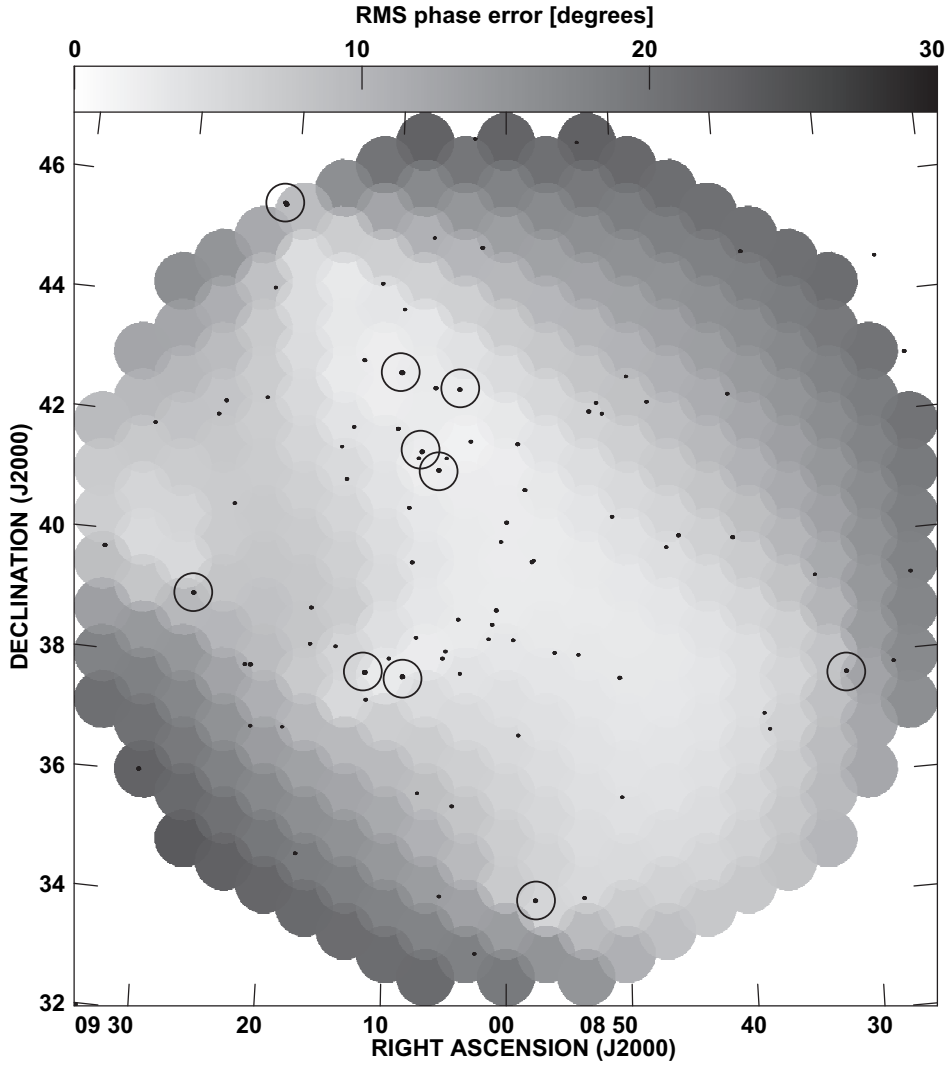
Figure 6.6: *The grayscale map represents the residual phase RMS between the distorting and correcting ionospheric phase models across the primary beam area, averaged over baselines and time. The phase RMS was calculated for a hexagonal grid of viewing directions across the FoV. Each viewing direction is depicted by a small circular area. Overplotted is a contour map of the point sources as seen in the SPAM image (which extends slightly beyond the grid of circles). The 10 peeling sources are marked by circles. The correspondence between the models is largest near the calibrator sources and over a large part of the inner primary beam. The discrepancy is largest near the South-East and North-West borders, away from the calibrators.*

6.4.3   Background Noise

In this and the next sections, we revert to analyzing image properties for an indirect, relative comparison between the different calibration techniques. In the presence of residual phase errors, part of the image background noise level consists of residual sidelobes after CLEANing. The local sidelobe noise increases with both the RMS phase error and the local source flux. When measured over a large image area, the mean sidelobe noise depends mainly on RMS phase error. For all relevant output images, the mean image noise σ was determined by fitting a Gaussian to the histogram of image pixel values from the inner quarter radius of the FoV (AIPS task IMEAN). Note that these images have not been corrected for primary beam attenuation. The results are given in Table 6.2.

Because no noise was added to the simulated J0900+398 data set, the resulting image noise of 3.0 mJy beam^{-1} in the undisturbed image is caused by incomplete UV coverage and inaccuracies in the imaging process (see Section 6.3.2), limiting the dynamic range to $\sim 10^4$. The local noise is highest near the sources, but significantly less near the brightest 10 sources with dedicated facets centered on their peak position. The SC and SPAM images from this data set were created using the same facet configuration. The SC image noise of 10.2 mJy beam^{-1} is 3.4 times as high as the undisturbed image noise, therefore dominated by phase error induced sidelobe noise. The SPAM image noise of 6.7 mJy beam^{-1} is a significant improvement over the SC image, but still 2.2 times as high as in the undisturbed image. The local noise in the SC and SPAM images has increased most apparently near bright sources as compared to the undisturbed image, which confirms the presence of residual phase errors after calibration.

For the real J0900+398 data set, both the SC and SPAM images have an image noise of ~ 70 mJy beam^{-1}. The SPAM image noise is slightly lower than SC. The local noise in the SC image is higher near bright sources. This is not the case in the SPAM image, which must be a direct result of an improved calibration accuracy near these sources. The FBC image noise for this data set is ~ 20 percent higher, a combination of a higher average noise over the FoV and higher local noise near bright sources.

For the real J1300-208 data set, the SPAM image has the same image noise as for the real J0900+398 data set, with no apparent increase near bright sources. At the same time, the noise levels in the SC and FBC images have increased with 30 and 35 percent, respectively. The noise in the SC image is highest near the bright sources. The FBC noise is highest near the brightest source and remains high in the rest of the image. The significant increase of the average FBC noise level indicates a dependence on ionospheric conditions, and therefore on calibration accuracy. The SPAM image noise appears to have little or no dependence on varying ionospheric conditions (Figure 6.7).

6.4.4   Source Properties

The presence of residual phase errors changes the apparent distribution of flux of a source (see Section 6.2.2). In the time-averaged image, sources may appear offset from their intrinsic position, may suffer from smearing or deformation, and sidelobes may be misidentified as sources. Comparing the properties of the same sources in differently calibrated

Field name	VLSS J0900+398 (simulated)	VLSS J0900+398 (real)	VLSS J1300-208 (real)
Mean background noise σ [mJy beam^{-1}]:			
Undisturbed	3.0	–	–
SC	10.2	71	92
FBC	–	87	118
SPAM	6.7	67	68
Number of sources with a peak flux larger than 5σ:			
Undisturbed	91	–	–
SC	91	393	374
FBC	–	310	285
SPAM	91	372	392
5σ source fraction with an NVSS counterpart within 80″:			
Undisturbed	1.	–	–
SC	1.	0.83	0.60
FBC	–	0.86	0.74
SPAM	1.	0.97	0.97

Table 6.2: *Overview of results from calibrating and imaging three test case data sets with no ionosphere (Undisturbed), self-calibration (SC), field-based calibration (FBC) and SPAM.*
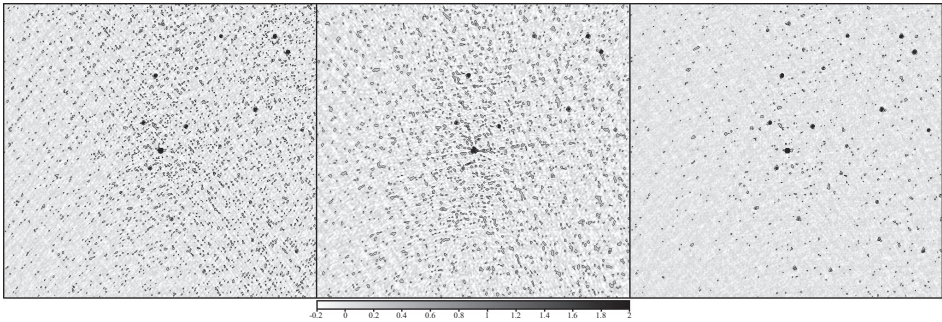
Figure 6.7: *Greyscale plots of a 3.5×3.5 square degree area in the VLSS J1300-208 field centered on the bright (40 Jy) point source 3C 283. All three images have contours (black lines) overplotted at [0.15, 0.48, 0.83, 1.16, 1.50] Jy. Left: Image after self-calibration, middle: image after field-based calibration, and right: image after SPAM calibration.*

images allows for a relative comparison of the performance of the different calibration techniques.

To allow for comparison of source properties, we applied the source extraction tool BDSM [90] on all relevant images. BDSM performed a multiple 2-dimensional Gaussian fit on islands of adjacent pixels with amplitudes above a specified threshold based on the *local* image noise σ_L in the image. Multiple overlapping Gaussians were grouped together into single sources. We applied BDSM to all images, using the default extraction criteria, except for the following: a source detection requires at least 4 adjacent pixel values above $2.5\,\sigma_L$, with at least one pixel value above $4\,\sigma_L$.

Source Counts

Due to the non-Gaussian character of the phase-induced sidelobe noise, the source catalogs will contain spurious detections. To suppress these, we removed sources with a peak flux smaller than $5\,\sigma$ from the catalogs. The remaining number of catalog entries are listed in Table 6.2. Additionally, each catalog was cross-associated against the NVSS catalog, which has a slightly higher resolution (45″). For an average spectral index of -0.8, the NVSS detection limit is at least 10 times lower than for the VLSS. At the risk of missing an incidental ultra-steep spectrum source, we determined the source fraction that has an NVSS counterpart within an 80″ radius (one VLSS beamsize), which are also listed in Table 6.2.

For the simulated J0900+398 data set, all 91 input sources are detected and matched against NVSS counterparts, regardless of the calibration method. Due to the low noise levels and the lower limit of 1 Jy on the input source catalog, all sources are effectively $\gtrsim 100\,\sigma$ detections. None of the sources had more than one Gaussian fitted to it, despite the freedom to do so.

For the real J0900+398 data set, the higher σ in the FBC image is reflected in a smaller number of source detections as compared to SC and SPAM. SC detects slightly more sources than SPAM, despite the slightly higher σ. However, there is a very large fraction of sources in the SPAM catalog that has an NVSS counterpart, significantly larger than for both the SC and FBC catalogs. This suggests that the SPAM catalog is much less contaminated by false detections than the SC and FBC catalogs, resulting in a larger absolute number of true detections.

This is further strenghtened by the results from the real J1300-208 data set. For this test case, the SPAM image has the largest number of source detections. Again, the SPAM catalog has the largest fraction of associations with the NVSS catalog, the same fraction as with the J0900+398 data set. In contrast, the fraction of NVSS counterparts for SC and VLSS have both gone down. This is best explained by an increase in (non-Gaussian) sidelobe noise in the image background due to calibration errors, which corresponds with the observed increase in σ.

Source Peak Fluxes

The presence of residual phase errors after calibration can cause an unresolved source shape to deviate from a point source shape. The source flux is redistributed over a larger area and the peak flux of the source drops. At $80''$ resolution, most sources in a VLSS field are unresolved. Therefore, a mean increase of source widths over the point source width is a direct measure of ionospheric conditions. This argument was used in the pre-selection of data sets for our test cases.

For significant source deformations or low SNR sources, determination of the shape of individual sources is subject to large uncertainties (e.g., Condon et al. [91]). Because determination of peak fluxes is much more robust, we use these for a relative comparison of calibration accuracy. Starting with the original catalogs as produced by BDSM, we associate sources between the undisturbed, FBC, SC and SPAM catalogs that lie within $80''$ of the same NVSS source and has a peak flux larger than $5\,\sigma$ in at least one of the two catalogs.

For the simulated J0900+398 data set, the true peak fluxes of all 91 sources are known. A comparison between peak fluxes from the undisturbed image and the input catalog identifies a small (< 1 percent) CLEAN bias of 3.6 mJy beam^{-1} [e.g. 77, 78, 92]. Ignoring the image noise dependency of CLEAN bias, we applied this small correction to the peak fluxes in the undisturbed, SC and SPAM source catalogs before proceeding. Figure 6.8 shows a comparison of the measured-to-input peak flux ratios for sources in the SC and SPAM images. The mean peak flux ratio for both images is approximately equal and just slightly smaller than one. The larger scatter in the SC peak fluxes is consistent with a higher σ. Using Equation 6.6, the random part of the mean RMS phase error for both SC and SPAM is estimated at 5–6 degrees. This value is comparable to the observed RMS phase error over large parts of the SPAM image (Section 6.4.2).

To study the nature of residual RMS phase errors after application of SPAM, we plot the RMS phase errors at the source positions from Figure 6.6 against SPAM-to-input peak flux ratios (Figure 6.9). For Gaussian random phase errors, the peak flux ratio
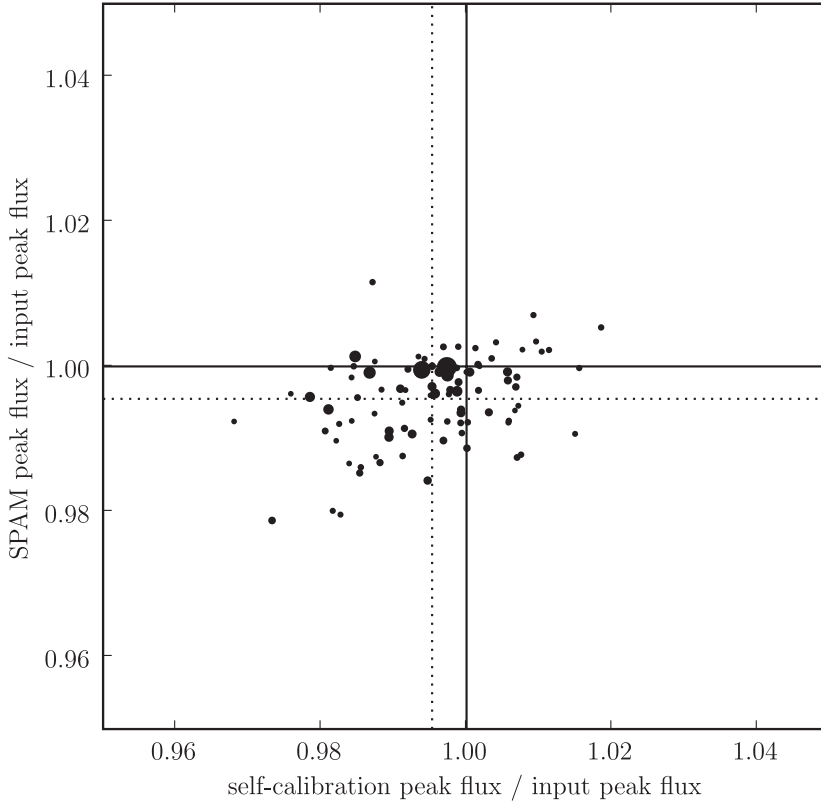
Figure 6.8: *Peak flux ratios of point sources in the simulated J0900+398 field. Peak fluxes were measured in the self-calibration image and the SPAM image, corrected for a small CLEAN bias and divided by the input model peak fluxes. The size of each dot scales with the input model peak flux, ranging from 1.02 to 26.7 Jy. Ideally (without phase errors), the peak flux ratios would be scattered around one (solid lines) due to image noise dependent errors in the peak flux determination. Instead, the peak flux ratio distributions along the x- and y-axis are centered around 0.995 and 0.996, respectively (dotted lines), which is a direct result of the residual phase errors. The smaller and larger scatter in distribution of SPAM- and self-calibration peak flux ratios is consistent with peak flux determination inaccuracies due image background noise levels.*
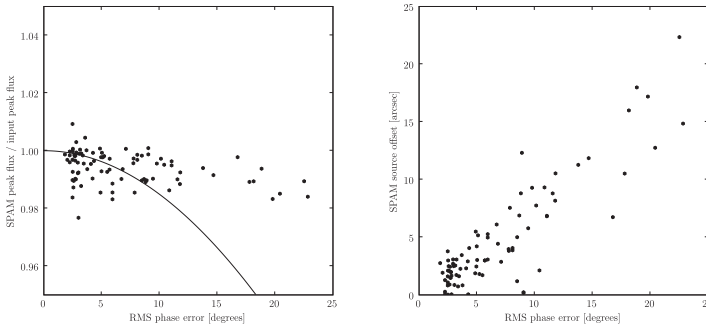
Figure 6.9: *Peak flux ratios in the simulated J0900+398 field: Left: Peak flux ratios of the 91 extracted sources from the SPAM image as compared to the input model sources, plotted as a function of the residual RMS phase error after SPAM calibration. Overplotted is the theoretical Strehl ratio (solid line) as given in Equation 6.6. For larger RMS phase errors, the measured peak flux ratios do not follow the theoretical strehl ratio curve. This indicates that systematic phase errors dominate the larger RMS phase errors. Right: Same peak flux ratios plotted as a function of absolute position offset between extracted sources in the SPAM image and the input model (see Figure 6.12). The presence of a strong correlation indicates that residual phase gradients dominate the larger RMS phase errors.*

is expected to decrease with increased RMS phase error as described in Equation 6.6. However, the discrepancy between the data points and Equation 6.6 indicates that for larger RMS values the phase errors are predominantly systematic rather than random.

For the real J0900+398 data set, Figure 6.10 shows a comparison of peak fluxes for associated sources in the SC, FBC and SPAM catalogs. There is a good match between peak fluxes measured in the SC and SPAM catalogs. For high SNR sources with a peak flux above 1 Jy, the SPAM peak fluxes match on average within 1 percent with the SC peak fluxes. Similarly, SC and SPAM peak fluxes are on average 10 percent higher than FBC peak fluxes. The systematic increase of peak fluxes for SC and SPAM as compared to FBC for many more than the calibrator sources denotes a more accurate calibration over large parts of the FoV. Towards the low flux end, source detections are slightly biased towards the image with the highest noise level, which is the FBC image.

Figure 6.11 shows the same comparison of peak fluxes for the real J1300-208 data set. For high SNR sources with a peak flux above 1 Jy, the SC peak fluxes are by far the smallest, while FBC and SPAM peak fluxes are on average higher by 15 and 24 percent, respectively. The relative loss of peak flux in the SC image is a clear indication of the break-down of the assumption of isoplanaticity across the FoV. Under the conditions that clearly need direction-dependent corrections, the SPAM peak fluxes are on average 7 percent higher than the FBC peak fluxes.
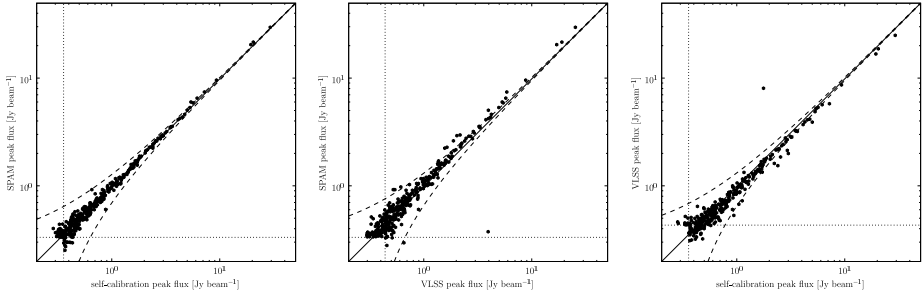
Figure 6.10: *Peak fluxes in the real J0900+398 field Left: Peak flux comparison for 367 sources detected in both the self-calibration and SPAM images. The straight diagonal line represents equality, the dashed lines represent $3\sigma_C$ deviations (where σ_C is the combined noise level from both images), and the dotted lines indicate the 5σ detection limit. For bright sources (peak fluxes $\gtrsim 1\,Jy\,beam^{-1}$), the average peak flux ratio is 1.00. Middle: Same for 329 sources in the field-based calibration (VLSS) and SPAM images. The average bright peak flux ratio of SPAM over field-based calibration is 1.10. Right: Same for 313 sources in the self-calibration and field-based calibration (VLSS) images. The average bright peak flux ratio of self-calibration over field-based calibration is 1.10. In all plots, the image noise causes a larger scatter in the peak flux determinations of faint sources ($\lesssim 1\,Jy\,beam^{-1}$) and consequently, a selection bias towards positively enhanced peak fluxes that increases with image noise.*
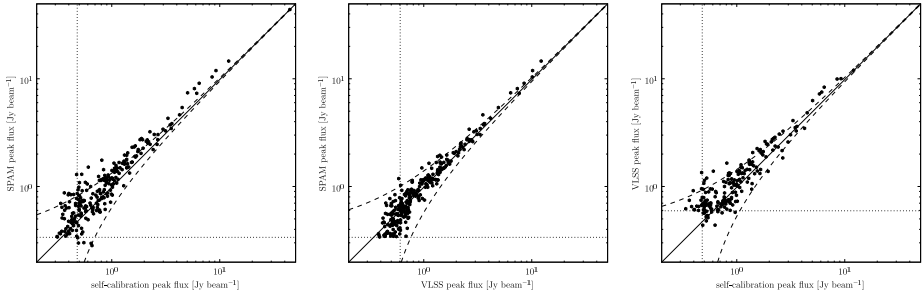

Figure 6.11: *Peak fluxes in the (real) J1300-208 field: Left: Peak flux comparison for 247 sources detected in both the self-calibration and SPAM images. For bright sources (peak fluxes $\gtrsim 1\,Jy\,beam^{-1}$), the average peak flux ratio of SPAM over SC is 1.24. Middle: Same for 278 sources in the field-based calibration (VLSS) and SPAM images. The average bright peak flux ratio of SPAM over field-based calibration is 1.07. Right: Same for 202 sources in the self-calibration and field-based calibration (VLSS) images. The average bright peak flux ratio of field-based calibration over self-calibration is 1.15.*
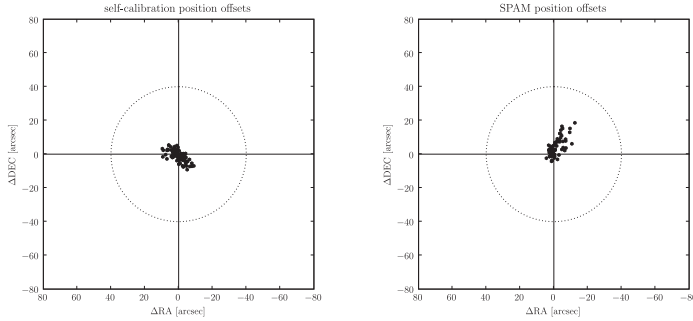
Figure 6.12: *Position offsets in the simulated J0900+398 field: Left: Offsets between the measured source positions in the self-calibration image as compared to the input model. Right: Same for the SPAM image. In both cases, the distribution around the origin is non-Gaussian. For the SPAM image, the tail of points extending roughly northwards indicates the presence of persistent phase gradients in local parts of the SPAM image. All source position offsets fall well within the size of the 80″ restoring beam (dotted line).*

Astrometry

When the time-average of residual phase errors towards a source contains a non-zero spatial gradient, the source will appear to have shifted its position in the final image (see Section 6.2.2). This gradient may indicate a limitation of the calibration model to reproduce the ionospheric phase corruptions (e.g., in the absence of nearby calibrators), but may also be introduced by the peeling process. The latter occurs when a peeling source is re-centered to the wrong catalog position (see Section 6.3.3). Because such an error propagates into the calibration model, many sources in the vicinity of the peeling source may also suffer from a systematic astrometric error.

For the simulated data set, the peak positions of sources as determined by BDSM were compared against the positions of counterparts in the input model. For the real data sets, we compared against the NVSS catalog instead. When comparing against NVSS positions, apparently large position offsets may occur due to resolution differences and spectral variation across the source. Averaged over a large number of sources, these offsets should have no preferential orientation. In contrast, a residual phase gradient in a certain viewing direction is expected to cause systematic offsets for groups of sources in a certain preferential direction.

For the simulated J0900+398 data set, Figure 6.12 shows that the positions for both SC and SPAM are accurate to within ~ 10″, except for a small tail of ~ 15 SPAM sources that have somewhat larger offsets. These sources are all positioned near the edge of the FoV, where the RMS phase error is large (Figure 6.6). Figure 6.9 also confirms this by the clear correlation between RMS phase error and absolute position offsets.

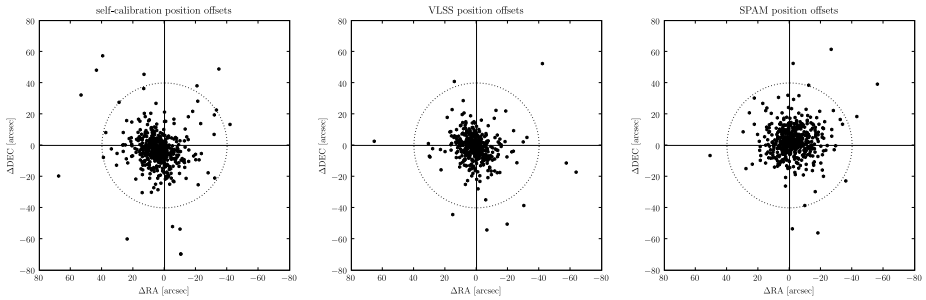For the real J0900+398 data set, the source position offsets for SC, FBC and SPAM

Figure 6.13: *Position offsets in the real J0900+398 field Left: Offsets between the measured source positions in the self-calibration image as compared to the NVSS catalog. Middle: Same for the field-based calibration (VLSS) image. Right: Same for the SPAM image.*

relative to NVSS catalog positions are plotted in Figure 6.13. The larger scatter as compared to the simulated J0900+398 data set can be the (combined) result of less accurate position measurements due to higher image noise, resolution and spectral differences between the observations and the NVSS catalog or larger residual RMS phase errors after calibration. The observed scatter for SC is centered around a point that is offset from the origin by $\sim 5''$, which is either caused by inaccuracies in the initial sky model or during the self-calibration process (Section 6.3.2). The scatter of both FBC and SPAM offsets is centered close to the origin. The RMS of the scatter around the mean position offset is $10.5''$ for both FBC and SPAM (despite the apparently larger scatter for SPAM, which is due to a larger number of data points), both smaller than the $11.9''$ for SC.

For the real J1300-208 data set, the source position offsets for SC, FBC and SPAM relative to NVSS catalog positions are plotted in Figure 6.14. The position scatter for all three methods is significantly larger than for the real J0900+398 data set, and all suffer from systematic position offsets in varying degrees of severity. The position offsets in the SC image have a seriously distorted distribution, which includes a large tail of points that extends roughly southwards. This indicates the presence of varying systematic source offsets over the whole FoV. The distribution of position offsets in the FBC image is more compact but also asymmetric, and is approximately centered around a point that is $\sim 10''$ offset in northward direction from the origin. A large number of the SPAM position offsets are clustered near the origin, similar to the real J0900+398 data set, but there is an additional tail of points that runs roughly northwards. The RMS of the scatter around the mean position offset is $20.7''$, $16.5''$ and $14.8''$ for SC, FBC and SPAM, respectively, which confirms the apparently strongest clustering of points in the SPAM position offset plot.

Systematic position offsets in the images can be reduced by distortion and regridding of the images. To this purpose, Cohen et al. [41] fit a fourth order Zernike polynomial to the (time constant) position offsets of typically more than 100 sources in the FBC images of the VLSS. They estimate that, after correction, the final residual position error in the
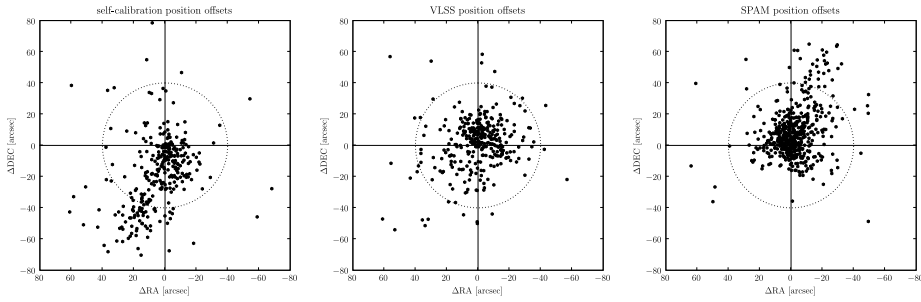
Figure 6.14: *Position offsets in the real J1300-208 field: Left: Offsets between the measured source positions in the self-calibration image as compared to the NVSS catalog. Middle: Same for the field-based calibration (VLSS) image. Right: Same for the SPAM image.*

full VLSS catalog due to the ionosphere is $\lesssim 3''$ in both RA and DEC.

6.5 Discussion and Conclusions

The SPAM method for ionospheric calibration has been succesfully tested on one simulated and two carefully selected visibility data sets of 74 MHz observations with the VLA (taken from the VLSS; Cohen et al. [41]). From the results of these test cases, we draw the following conclusions:

(i) A proof-of-concept is given for several different techniques that were incorporated in SPAM calibration. The peeling technique [26] was succesful in providing relative measurements of ionospheric phase errors in the direction of several bright sources in the FoV. The Karhunen-Loève phase screen (Chapter 5 [see also 51]) at fixed height was able to combine these measurements into a consistent model per time stamp. For relatively bad ionospheric conditions, it was demonstrated that the ionospheric calibration cycle (repeated ionospheric calibration and subsequent imaging; Noordam [26]) converges within a few iterations to a calibration of similar accuracy as under relatively good ionospheric conditions (for which one iteration was sufficient).

(ii) Ionospheric calibration with SPAM is more accurate than the existing self-calibration [e.g. 11] and field-based calibration [42] techniques. Even for relatively compact array configuration like VLA-B and BnA, significant improvements in image quality are obtained by allowing for higher-order (i.e., more than a gradient) spatial phase corrections over the array in any viewing direction. In the resulting images, we obtained dynamic range improvements of 5–45% and 70–80% under relatively good and bad ionospheric conditions, respectively.

(iii) Although the mean astrometric accuracy of source positions in SPAM images is similar to or better than for self-calibration and field-based calibration, systematically larger astrometric errors are present in regions of the output images of all calibration meth-

ods. This is caused by a shortage of available calibrators in these regions and positional inaccuracies in the reference source catalog used for calibration.

The 65 mJy beam^{-1} noise levels in the SPAM images match the lowest noise levels of the more than 500 images that define the VLSS survey. A potential reduction of the average noise level from 100 mJy beam^{-1} to 65 mJy beam^{-1} for the full VLSS survey would significantly increase the number of source detections from ~70,000 to about 120,000 (an increase of ~ 75%), but also it would greatly enhance virtually every science goal. For example, using the radio luminosity function for high-luminosity radio galaxies from Jarvis et al. [93], the estimated number of detectable HzRGs in the VLSS would increase by 65%, but also the maximum redshift would increase. For a luminous radio galaxy with luminosity of 2×10^{28} W Hz^{-1} sr^{-1} at 74 MHz, the redshift limit would rise from $z = 5.7$ to $z = 6.8$. Another example is the detection and study of cluster radio halos. Using available halo population models [94, 95], the anticipated noise reduction would roughly double the number of detectable halo systems.

For the VLSS, the estimated theoretical thermal noise level of 35 mJy beam^{-1} is still a factor of two lower than the average background noise level of ~ 65 mJy beam^{-1} in the SPAM images. From inspection of the SPAM images we cannot identify an obvious single cause for this. Therefore, similar to Cohen et al. [41], we expect the remaining excess noise to be the combined result of several different causes, including residual ionospheric phase errors after SPAM calibration, but also residual RFI, collective sidelobe noise from many non-deconvolved sources (too faint or outside the FoV) and variable source amplitude errors (e.g., due to pointing errors and non-circular antenna beam patterns; see Bhatnagar et al. [88]).

The SPAM test results indicate that the ionospheric calibration accuracy may be further improved. The typical model fit RMS phase error per antenna of ~ 20 − 30 degrees for real data sets is much larger than the 3 degrees for the noiseless simulated data set. There are several possible sources of error, either in the peeling phase corrections or the ionospheric phase model. Noise in the visibilities (either thermal or non-thermal), contamination from other sources, inaccuracies in the peeling source model and undersampling of the fastest phase fluctuations are factors that degrade the accuracy of peeling. Also, the ionospheric phase screen model may be a poor representation of reality, either because it is incomplete (e.g., absence of vertical structure) or the fixed model parameters are chosen poorly (e.g., screen height, spectral index of phase fluctuations). Several of these issues will be addressed in future work (Section 6.6).

The potential problems with the peeling technique raises the question whether one should use alternative methods. Apart from the precautions described in Section 6.3.3, we have found little means to improve the accuracy of the peeling process for single sources any further. One unexplored option is to peel sources in groups, e.g. identify isoplanatic patches of sky with a large enough total flux from multiple sources. Two possible alternatives approaches to peeling are: (i) simultaneous self-calibration towards multiple sources in the FoV, or (ii) fitting the ionosphere model directly to the visibilities rather than using peeling as an intermediate step. Although these alternative approaches have not been tested by us in practise, we anticipate little improvement over our current accuracy. In Chapter 3 [see also 23] it was shown, by simulation, that iterative peel-

ing converges to the same solution as simultaneous self-calibration. A direct fit of the ionosphere model to the visibilities is, similar to self-calibration, biased towards accurate solving in the direction of the apparently strongest source in the FoV. Although not conclusive for this approach, tests with SPAM show that using even a moderate flux-based weighting into the ionospheric phase model fitting against peeling phase corrections introduces a large bias towards the brightest source, while calibration accuracy towards other peeled sources degrades severely.

For the existing and future large low-frequency radio interferometer arrays like VLA-A, GMRT, LOFAR, LWA and SKA, the need for a direction-dependent ionospheric calibration method is evident. Based on the results presented in this paper, it is difficult to draw quantitative conclusions on the achievable calibration accuracy for these arrays. If a SPAM-like calibration algorithm is to be used in a very high signal-to-noise observing regime under quiet to moderate ionospheric conditions, it seems likely that residual RMS phase errors in the order of a few degrees could be achieved, comparable to the SPAM results on the simulated VLSS data set.

When relying on the array itself to provide the necessary measurements to constrain ionospheric correction models, ionospheric calibration requires an array layout and sensitivity that allows for sampling the ionsphere over the array at the relevant spatial scales and time resolution. The spatial sampling is determined by the instantaneous pierce point distribution (or more general, the distribution of lines-of-sight through the ionosphere), which depends on the array layout and the detectable calibrator constellation. For future design of low-frequency arrays, it is recommended to optimize the array layout not just for scientific arguments (in general, centrally dense and sparse outside for good UV coverage), but also for ionospheric calibrability (in general, both uniform and randomized).

6.6   Future Work

To test the robustness and limitations of the method, it is necessary to apply SPAM calibration on a wide variety of data sets at different (low) frequencies, obtained with different arrays under different ionospheric conditions. Our highest priority is to test SPAM on observations from the largest existing LF arrays; the VLA in A-configuration and the GMRT. Data for these tests have been obtained and tests are currently in progress. One important possible limitation is the use of a 2-dimensional phase screen to represent the ionosphere. We plan to expand the SPAM model by including multiple screens at different heights and compare the resulting image properties against the current single screen model.

Another limitation of the current implementation is the absence of restrictions on the time behaviour of the model. Antenna-based peeling phases clearly show a coherent temporal behaviour, which is likely to exist for physical reasons. This could be used to reduce the number of required model parameters and suppress the noise propagation from the peeling solutions. We are currently investigating the possibilities of forcing the SPAM model to be continuous in time. In chapter 7 we will present a few ideas on how this can be done without adding much computational complexity.

The MAP estimator was not fully implemented in SPAM. The phase screen was composed of the Karhunen-Loève basis function, but the term representing the a priori information was not included in the least squares cost function. The main reason for not implementing this was that the ionospheric conditions, i.e. the slope and scaling of the powerlaw, were not known in advance. In Chapter 5 is described how these hyperparameters can be estimated from the data. Once this has been implemented, the number of basis functions can be increased without the risk of lower performance due to noisier estimates.

Several of the authors of the article om which this chapter is based are currently involved in setting up a simulation framework in which one has full control over the sky emission, ionospheric behaviour and array characteristics when generating artificial low frequency observations. Like in the test case on simulated data presented in Section 6.4, this allows for direct and quantitative comparison between the distorting ionosphere model and the recovered ionospheric phase model by SPAM. We plan to use this setup to further test optimize SPAM calibration for a broad range of ionospheric conditions.

Chapter 7

Extensions

Bayesian estimation is a generic method with a wide range of applications. In this thesis we have applied the method to a single layer model of the ionosphere with a power law distribution of the electron density fluctuations. The method has been proved to be relatively easy to implement, robust and to have a better performance than existing methods based on deterministic models. In this chapter we will explore a few possible applications of Bayesian estimation to more extended data models. Most of these extensions have not been tried in practice yet, but based on the successful application of Bayesian estimation described in the previous chapter we are hopeful that in practice at least some these extensions will perform better than estimators based on deterministic models.

7.1 Introduction

The starting point for the extensions will be the MAP estimator as described in section 5.3.3. The underlying data model of the observed data $\mathbf{x}$ is assumed to be of the form

$$\mathbf{x} = \mathbf{f}(\theta) + \mathbf{w}. \tag{7.1}$$

The function $\mathbf{f}$ describes the relationship between the unknown parameters θ and the expected value of the data. The function $\mathbf{f}$ is known as the Measurement Equation in radio astronomical literature. The parameter vector θ contains all unknowns, such as the ionopshere induced phase shifts, instrument gain. The actually observed data is contaminated by noise, modeled as the noise vector $\mathbf{w}$. The data is the output of the correlator, so $\mathbf{x}$ contains the entries of correlation matrix estimates. The correlation matrix estimates have a Wishart distribution. For simplicity the distribution of $\mathbf{w}$ is assumed to be Gaussian, with covariance matrix $\mathbf{C_w}$ and zero mean. The covariance of the noise vector can be estimated from the data by using the well known expression for the covariance of Wishart distributed data [96]

$$\mathbf{C_w} = \frac{1}{N}\bar{\mathbf{R}} \otimes \mathbf{R}, \tag{7.2}$$
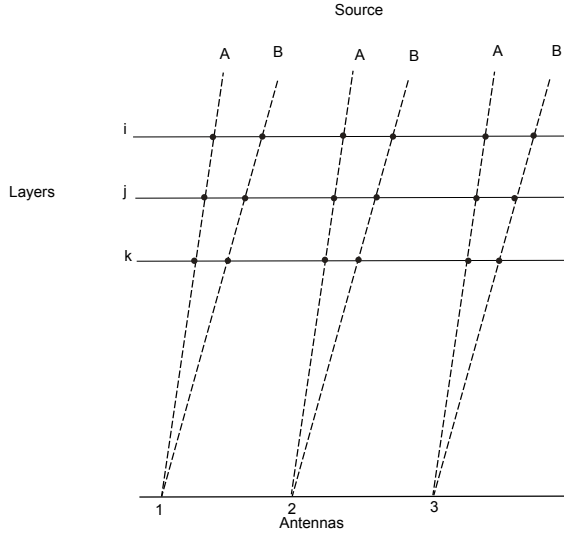
Figure 7.1: *Multilayer model*

where $\mathbf{R}$ is the expected value of the Wishart distributed data. The entries of $\mathbf{x}$ are estimates of the entries of $\mathbf{R}$. By substiting $\mathbf{R}$ with the estimate of $\mathbf{R}$, an estimate of $\mathbf{C_w}$ can be found. The a priori information is given in the form of an expected value $\bar{\theta}$ and covariance matrix $\mathbf{C_\theta}$. Again the distribution is assumed to be Gaussian.

The general MAP estimator solves the following minimization problem

$$\hat{\theta} = \arg\min_{\theta}(\mathbf{f}(\theta) - \mathbf{x})^{\mathrm{H}}\mathbf{C_w}^{-1}(\mathbf{f}(\theta) - \mathbf{x}) + (\theta - \bar{\theta})^{\mathrm{T}}\mathbf{C_\theta}^{-1}(\theta - \bar{\theta}) \tag{7.3}$$

In the next sections we will present a few ideas on how we could find $\mathbf{C_\theta}$ for ionospheric models that go beyond a single layer. We also present some ideas on how (7.3) can be solved efficiently when correlations over time are included and θ grows very large.

7.2   Multilayer model

Consider the ionospheric model again, where the vector θ contains the unknown ionospheric phases. In chapters 5 and 6 this problem was solved by using a single layer model assuming that the electron density fluctuations are confined to a thin layer. In reality the fluctuations are distributed over height. A relatively simple way to capture the distribution over height is by using multiple independent layers at different heights, as shown in Figure 7.1. Let the ionospheric phase for the kth layer be given by the vector θ_k. The total ionospheric phase is the sum of the layers:

$$\theta = \sum_{k=1}^{K}\theta_k,$$

where K is the number of layers. Because the layers are independent, the correlation matrix $\mathbf{C}_{\boldsymbol{\theta}}$ is sum of the correlation matrices for each layer:

$$\mathbf{C}_{\boldsymbol{\theta}} = \sum_{k=1}^{K} \mathbf{C}_{\boldsymbol{\theta}_k}.$$

The matrices $\mathbf{C}_{\boldsymbol{\theta}_k}$ are computed based on a single layer model as given in equation (5.6). This matrix can then be substituted in the standard MAP equation (7.3).

The interpolated values for directions other than the calibrator sources are also summations over the layers

$$\rho = \sum_{k=1}^{K} \rho_k,$$

where ρ and ρ_k are the ionospheric phases at the interpolation point, respectively for the total line of sight and the individual layers. The estimates are given according to Kriging's interpolation equation (see section 5.3.4)

$$\hat{\rho} = \mathbf{C}_{\rho\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}, \tag{7.4}$$

where

$$\mathbf{C}_{\rho\boldsymbol{\theta}} = \sum_{k=1}^{K} \mathbf{C}_{\rho_k\boldsymbol{\theta}_k} \tag{7.5}$$

Again the covariance matrix $\mathbf{C}_{\rho_k\boldsymbol{\theta}_k}$ is by the equation for the single layer model (5.6) . Kriging interpolation can also be applied to the layers

$$\hat{\rho}_k = \mathbf{C}_{\rho_k\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}}^{-1}\hat{\boldsymbol{\theta}}. \tag{7.6}$$

From equations (7.4), (7.5) and (7.6) it follows that the Kriging estimate of the sum of the layers is the sum of the Kriging estimates of the layers

$$\hat{\rho} = \sum_{k=1}^{K} \hat{\rho}_k.$$

The difficulty with this approach is that one needs to know how many layers to use, at what height these layers need to be placed and what power law (β and s_0) to use for each of the layers. Since a radio telescope sees only the aggreagate phase, not the individual layers, there is no straighforward way the estimate these parameters from the data. Tomography, physical models and external observations might be necessary to estimate these parameters.

7.3 Full 3D Model

For a three dimensional model of the ionosphere the electron density is described by the function $e(x, y, z)$. The ionospheric phase is given by an integral over the line of sight
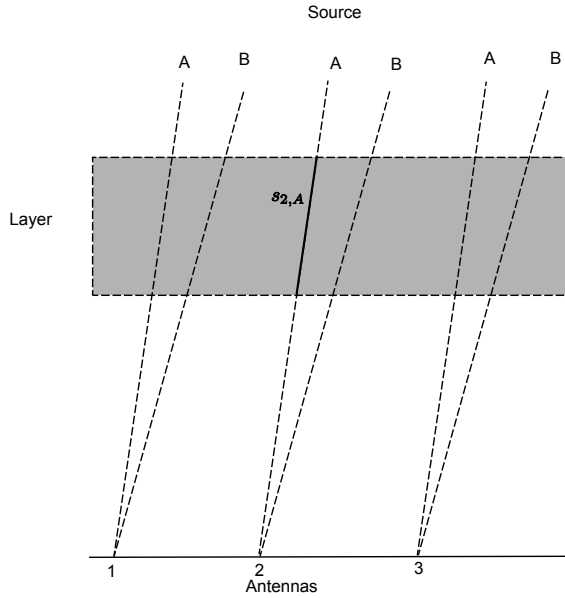
Figure 7.2: *Three dimensional model. The solid line marked by $s_{2,A}$ is the path through the ionospheric layer from Antenna 2 to Source B. The integral of the electron density over the path yields the total electron content between the antenna and the source.*

(see Figure 7.2)

$$\theta_k = C\lambda \int_{s_k} e(x, y, z) ds_k.$$

where s_k is the path through the ionosphere along the line of sight, λ the wavelength and C a constant. The correlations between two ionospheric phases is a double integral over the two lines of sight s_k and s_l

$$C_{\theta_k\theta_l} = C^2\lambda^2 \int_{s_k} \int_{s_l} C_\phi(x_k, y_k, z_k, x_l, y_l, z_l) ds_l ds_k$$

For most models the integrals need to be evaluated numerically. This could also be described as a many layer model similar to the multi layer model described in the previous section. The difference between these two models is that the layers in the many layer model are not independent.

The difficulties here are similar to the multilayer model, but more complicated because also the correlation between diffent heights needs to be known.

7.4   Temporal model

The correlator generates new output for each time block. From here on the subscript k will be used to distinguish between the different blocks in time, for instance $\mathbf{x}_k$, $\mathbf{f}_k$, $\boldsymbol{\theta}_k$ and $\mathbf{w}_k$. To refer to the set of all the blocks up to the kth block the index $\{k\}$ is used, for instance

$$
\mathbf{x}_{\{k\}} = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_k \end{bmatrix}, \quad \boldsymbol{\theta}_{\{k\}} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_k \end{bmatrix}, \quad \mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}}) = \begin{bmatrix} \mathbf{f}_1(\boldsymbol{\theta}_1) \\ \vdots \\ \mathbf{f}_k(\boldsymbol{\theta}_k) \end{bmatrix}
$$

Note that function $\mathbf{f}_k(\boldsymbol{\theta}_k)$ only depends on parameters $\boldsymbol{\theta}_k$, not on the parameters corresponding to the other time blocks. The only way the different time blocks are tied together is through the correlation matrix $\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}$. The function $\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}})$ establishes no connection between the different time blocks. The correlation matrix $\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}$ can be partitioned in blocks

$$
\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}} = \begin{bmatrix} \mathbf{C}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_1} & \dots & \mathbf{C}_{\boldsymbol{\theta}_1 \boldsymbol{\theta}_k} \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{\boldsymbol{\theta}_k \boldsymbol{\theta}_1} & \dots & \mathbf{C}_{\boldsymbol{\theta}_k \boldsymbol{\theta}_k} \end{bmatrix}.
$$

Matrix $\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}$ is a full matrix, because the parameters are asssumed to vary slowly over time and hence are correlated over time. This means that the off diagonal blocks are non-zero. The correlation over time needs to be estimated from the data or derived from theoretical considerations. This is still an open problem. When LOFAR is operational it will generate a lot of data on the ionosphere that can be used to estimate the correlation and verify theoretical models.

The noise is assumed to be uncorrelated over time, so $\mathbf{C}_\mathbf{w}$ is a block diagonal matrix

$$
\mathbf{C}_{\mathbf{w}_{\{k\}}} = \begin{bmatrix} \mathbf{C}_{\mathbf{w}_1} & & \\ & \ddots & \\ & & \mathbf{C}_{\mathbf{w}_k} \end{bmatrix}.
$$

7.4.1   Sequential Estimation

In principle the large vectors and covariance matrices can be substituted into the general MAP estimator (7.3) yielding

$$
\hat{\boldsymbol{\theta}}_{\{k\}} = \arg \min_{\boldsymbol{\theta}_{\{k\}}} \left(\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}}) - \mathbf{x}_{\{k\}} \right)^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} \left(\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}}) - \mathbf{x}_{\{k\}} \right) +
$$

$$
\left(\boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \right)^{\mathrm{T}} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1} \left(\boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \right). \tag{7.7}
$$

In practice this problem is too large to fit in memory. Partitioning of the problem is needed to solve it. The partioning would be easy when the parameters were uncorrelated over time. In that case $\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}$ would be a block diagonal matrix and the minimization

problem could be solved for each k independently

$$\hat{\theta}_k = \arg\min_{\theta_k} (\mathbf{f}_k(\theta_k) - \mathbf{x}_k)^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_k}^{-1} (\mathbf{f}_k(\theta_k) - \mathbf{x}_k) +$$
$$\left(\theta_k - \bar{\theta}_k \right)^{\mathrm{T}} \mathbf{C}_{\theta_k}^{-1} \left(\theta_k - \bar{\theta}_k \right). \tag{7.8}$$

This is far less time consuming than solving the joint problem. Ignoring the correlation over time however gives poorer estimates. In this section a method will be described that finds an approximate solution to the joint problem by solving only problems of the size of a single time block.

Suppose the estimate $\hat{\theta}_{\{k\}}$ has been obtained based on $\mathbf{x}_{\{k\}}$. Now new data $\mathbf{x}_{k+1}$ becomes available. The joint problem is given by

$$\hat{\theta}_{\{k+1\}} = \arg\min_{\theta_{\{k+1\}}} (\mathbf{f}_{\{k+1\}}(\theta_{\{k+1\}}) - \mathbf{x}_{\{k+1\}})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k+1\}}}^{-1} (\mathbf{f}_{\{k+1\}}(\theta_{\{k+1\}}) - \mathbf{x}_{\{k+1\}}) +$$
$$\left(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}} \right)^{\mathrm{T}} \mathbf{C}_{\theta_{\{k+1\}}}^{-1} \left(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}} \right) \tag{7.9}$$

The previously found estimate $\hat{\theta}_{\{k\}}$ can be used as an initial point. A good initial point will reduce the number of steps the solver needs, but still the computational cost for each step is large because the number of free parameters and data points is large. A large computational saving can be made by keeping $\theta_{\{k\}}$ fixed and solving only for θ_{k+1}. This will lead to degraded performance compared to joined optimization since only information from the past is used. Future time blocks will not change the current result. However, this is an improvement over separately solving for each time block. The problem now becomes

$$\hat{\theta}_{k+1} = \arg\min_{\theta_{k+1}} (\mathbf{f}_{\{k+1\}}(\theta_{\{k+1\}}) - \mathbf{x}_{\{k+1\}})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k+1\}}}^{-1} (\mathbf{f}_{\{k+1\}}(\theta_{\{k+1\}}) - \mathbf{x}_{\{k+1\}}) +$$
$$\left(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}} \right)^{\mathrm{T}} \mathbf{C}_{\theta_{\{k+1\}}}^{-1} \left(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}} \right). \tag{7.10}$$

The constant terms in the minimization can be eliminated, yielding

$$\hat{\theta}_{k+1} = \arg\min_{\theta_{k+1}} (\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{k+1}}^{-1} (\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$
$$\left(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}} \right)^{\mathrm{T}} \mathbf{C}_{\theta_{\{k+1\}}}^{-1} \left(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}} \right). \tag{7.11}$$

A further simplification can be made by partioning the last term of the equation above in blocks

$$\hat{\theta}_{k+1} = \arg\min_{\theta_{k+1}} (\mathbf{f}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{k+1}}^{-1} (\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$
$$\begin{bmatrix} \left(\theta_{\{k\}} - \bar{\theta}_{\{k\}} \right)^{\mathrm{T}} & \left(\theta_{k+1} - \bar{\theta}_{k+1} \right)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{\theta_{\{k\}}\theta_{\{k\}}} & \mathbf{C}_{\theta_{\{k\}}\theta_{k+1}} \\ \mathbf{C}_{\theta_{k+1}\theta_{\{k\}}} & \mathbf{C}_{\theta_{k+1}\theta_{k+1}} \end{bmatrix}^{-1} \begin{bmatrix} \theta_{\{k\}} - \bar{\theta}_{\{k\}} \\ \theta_{k+1} - \bar{\theta}_{k+1} \end{bmatrix} \tag{7.12}$$

The matrix inverse in the last term of this minimization problem can be rewritten using the Schur complement as

$$
\begin{bmatrix} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}} & \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{k+1}} \\ \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}} & \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{k+1}} \end{bmatrix}^{-1} =
$$
$$
\begin{bmatrix} \mathbf{I} & -\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1}\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{k+1}} \\ 0 & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1} & 0 \\ 0 & \mathbf{S}^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1} & \mathbf{I} \end{bmatrix} \quad (7.13)
$$

where

$$
\mathbf{S} = \left(\mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{k+1}} - \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1}\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{k+1}} \right)
$$

Now define

$$
\tilde{\boldsymbol{\theta}}_{k+1} = \bar{\boldsymbol{\theta}}_{k+1} + \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1} \left(\boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \right).
$$

The vector $\tilde{\boldsymbol{\theta}}_{k+1}$ is a prediction of $\boldsymbol{\theta}_{k+1}$ based on $\boldsymbol{\theta}_{\{k\}}$. We can now write

$$
\begin{bmatrix} \boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \\ \boldsymbol{\theta}_{k+1} - \tilde{\boldsymbol{\theta}}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 \\ -\mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}}\mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \\ \boldsymbol{\theta}_{k+1} - \bar{\boldsymbol{\theta}}_{k+1} \end{bmatrix} \quad (7.14)
$$

Substituting equation (7.13) and then equation (7.14) into equation (7.12) leads to

$$
\hat{\boldsymbol{\theta}}_{k+1} = \arg\min_{\boldsymbol{\theta}_{k+1}} (\mathbf{f}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1}) +
$$
$$
\begin{bmatrix} \left(\boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \right)^{\mathrm{T}} & \left(\boldsymbol{\theta}_{k+1} - \tilde{\boldsymbol{\theta}}_{k+1} \right)^{\mathrm{T}} \end{bmatrix} \begin{bmatrix} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}\boldsymbol{\theta}_{\{k\}}}^{-1} & 0 \\ 0 & \mathbf{S}^{-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{\{k\}} - \bar{\boldsymbol{\theta}}_{\{k\}} \\ \boldsymbol{\theta}_{k+1} - \tilde{\boldsymbol{\theta}}_{k+1} \end{bmatrix} \quad (7.15)
$$

Eliminating the constant term results in the following minimization problem

$$
\hat{\boldsymbol{\theta}}_{k+1} = \arg\min_{\boldsymbol{\theta}_{k+1}} (\mathbf{f}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1}) +
$$
$$
\left(\boldsymbol{\theta}_{k+1} - \tilde{\boldsymbol{\theta}}_{k+1} \right)^{\mathrm{T}} \mathbf{S}^{-1} \left(\boldsymbol{\theta}_{k+1} - \tilde{\boldsymbol{\theta}}_{k+1} \right) \quad (7.16)
$$

This problem has the same size as estimating just block $k + 1$ only. The additional work is finding $\tilde{\boldsymbol{\theta}}_{k+1}$ and $\mathbf{S}^{-1}$.

7.4.2   Update Previous Estimates by Linear Approximation

In the previous section information from the past was used to estimate the parameters of the current time block, but the parameters from the past were not altered. In this section the parameters from the past will be updated. The functions $\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}})$ however will be replaced by their linearizations around previously found solutions $\boldsymbol{\theta}_{\{k\},0}$. The linearization for time block k given by

$$
\mathbf{f}_k(\boldsymbol{\theta}_k) \approx \mathbf{f}_k(\boldsymbol{\theta}_{k,0}) + \mathbf{J}_k(\boldsymbol{\theta}_{k,0})(\boldsymbol{\theta}_k - \boldsymbol{\theta}_{k,0}), \quad (7.17)
$$

where $\mathbf{J}_k(\boldsymbol{\theta}_{k,0})$ is the Jacobian of $\mathbf{f}_k(\boldsymbol{\theta}_k)$ at $\boldsymbol{\theta}_{k,0}$,

$$\mathbf{J}_k(\boldsymbol{\theta}_{k,0}) = \left. \frac{\partial \mathbf{f}_k(\boldsymbol{\theta}_k)}{\partial \boldsymbol{\theta}_k^{\mathrm{T}}} \right|_{\boldsymbol{\theta}_k = \boldsymbol{\theta}_{k,0}} . \tag{7.18}$$

The linearization for the set of time blocks $\{k\}$ is given by

$$\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}}) \approx \mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0}) + \mathbf{J}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0})(\boldsymbol{\theta}_{\{k\}} - \boldsymbol{\theta}_{\{k\},0}), \tag{7.19}$$

where $\mathbf{J}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0})$ is a block diagonal matrix,

$$\mathbf{J}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0})) = \begin{bmatrix} \mathbf{J}_1(\boldsymbol{\theta}_{1,0}) & & 0 \\ & \ddots & \\ 0 & & \mathbf{J}_k(\boldsymbol{\theta}_{k,0}) \end{bmatrix} \tag{7.20}$$

From now on we will simply write $\mathbf{J}_k$ and $\mathbf{J}_{\{k\}}$ for $\mathbf{J}_k(\boldsymbol{\theta}_{k,0})$ and $\mathbf{J}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0})$ respectively. Now let us start with the joint optimization problem for $k + 1$ data blocks

$$\hat{\boldsymbol{\theta}}_{\{k+1\}} = \arg \min_{\boldsymbol{\theta}_{\{k+1\}}} (\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}}) - \mathbf{x}_{\{k\}})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} (\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\}}) - \mathbf{x}_{\{k\}}) +$$
$$(\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{k+1}}^{-1} (\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1}) +$$
$$(\boldsymbol{\theta}_{\{k+1\}} - \bar{\boldsymbol{\theta}}_{\{k+1\}})^{\mathrm{T}} \mathbf{C}_{\boldsymbol{\theta}_{\{k+1\}}}^{-1} (\boldsymbol{\theta}_{\{k+1\}} - \bar{\boldsymbol{\theta}}_{\{k+1\}}) \tag{7.21}$$

Substituting approximation (7.19) into the equation above leads to

$$\hat{\boldsymbol{\theta}}_{\{k+1\}} = \arg \min_{\boldsymbol{\theta}_{\{k+1\}}} (\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0}) + \mathbf{J}_{\{k\}}(\boldsymbol{\theta}_{\{k\}} - \boldsymbol{\theta}_{\{k\},0}) - \mathbf{x}_{\{k\}})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} \times$$
$$(\mathbf{f}_{\{k\}}(\boldsymbol{\theta}_{\{k\},0}) + \mathbf{J}_{\{k\}}(\boldsymbol{\theta}_{\{k\}} - \boldsymbol{\theta}_{\{k\},0}) - \mathbf{x}_{\{k\}}) +$$
$$(\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{k+1}}^{-1} (\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1}) +$$
$$(\boldsymbol{\theta}_{\{k+1\}} - \bar{\boldsymbol{\theta}}_{\{k+1\}})^{\mathrm{T}} \mathbf{C}_{\boldsymbol{\theta}_{\{k+1\}}}^{-1} (\boldsymbol{\theta}_{\{k+1\}} - \bar{\boldsymbol{\theta}}_{\{k+1\}}) \tag{7.22}$$

Now we need the following basic result

Lemma 7.1. *In a minimization problem over $\boldsymbol{\theta}$ a term of the form*

$$(\mathbf{A}\boldsymbol{\theta} - \mathbf{r})^{\mathrm{T}} (\mathbf{A}\boldsymbol{\theta} - \mathbf{r}) \tag{7.23}$$

can be rewritten in the form

$$\left(\boldsymbol{\theta} - \check{\boldsymbol{\theta}}\right)^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{A} \left(\boldsymbol{\theta} - \check{\boldsymbol{\theta}}\right), \tag{7.24}$$

where

$$\check{\boldsymbol{\theta}} = (\mathbf{A}^{\mathrm{T}} \mathbf{A})^{-1} \mathbf{r} \tag{7.25}$$

Proof. Expanding (7.23) leads to

$$\boldsymbol{\theta}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{A} \boldsymbol{\theta} - \boldsymbol{\theta}^{\mathrm{T}} \mathbf{A}^{\mathrm{T}} \mathbf{r} - \mathbf{r}^{\mathrm{T}} \mathbf{A} \boldsymbol{\theta} + \mathbf{r}^{\mathrm{T}} \mathbf{r} \tag{7.26}$$

and expanding (7.25) leads to

$$\theta^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{A}\theta - \theta^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{r} - \mathbf{r}^{\mathrm{T}}\mathbf{A}\theta + \mathbf{r}\mathbf{A}^{\mathrm{T}}(\mathbf{A}^{\mathrm{T}}\mathbf{A})^{-1}\mathbf{A}^{\mathrm{T}}\mathbf{r} \tag{7.27}$$

The only difference between (7.26) and (7.27) is the last term, which is independent of θ hence for the purpose of minimization over θ (7.23) and (7.24) are equivalent.   □

The first term of (7.22) is of the form defined in Lemma 7.1, where

$$\mathbf{A} = \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1/2}\mathbf{J}_{\{k\}}, \tag{7.28}$$

$$\mathbf{r} = \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1/2}\left(\mathbf{x}_{\{k\}} - \mathbf{f}_{\{k\}}(\theta_{\{k\},0}) + \mathbf{J}_{\{k\}}\theta_{\{k\},0}\right). \tag{7.29}$$

$$\check{\theta}_{\{k\}} = \left(\mathbf{J}_{\{k\}}^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1}\mathbf{J}_{\{k\}}\right)^{-1}\mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1/2}\left(\mathbf{x}_{\{k\}} - \mathbf{f}_{\{k\}}(\theta_{\{k\},0}) + \mathbf{J}_{\{k\}}\theta_{\{k\},0}\right) \tag{7.30}$$

Equation (7.22) can be rewritten as

$$\hat{\theta}_{\{k+1\}} = \arg\min_{\theta_{\{k+1\}}}(\theta_{\{k\}} - \check{\theta}_{\{k\}})^{\mathrm{T}}\mathbf{J}_{\{k\}}^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1}\mathbf{J}_{\{k\}}(\theta_{\{k\}} - \check{\theta}_{\{k\}}) +$$
$$(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$
$$(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}})^{\mathrm{T}}\mathbf{C}_{\theta_{\{k+1\}}}^{-1}(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}}) \tag{7.31}$$

The linear term can be written as a norm,

$$\hat{\theta}_{\{k+1\}} = \arg\min_{\theta_{\{k\}}}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$
$$\left\|\left(\mathbf{J}_{\{k\}}^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1}\mathbf{J}_{\{k\}}\right)^{1/2}(\theta_{\{k\}} - \check{\theta}_{\{k\}})\right\|^2 +$$
$$\left\|\mathbf{C}_{\theta_{\{k+1\}}}^{-1/2}(\theta_{\{k+1\}} - \bar{\theta}_{\{k+1\}})\right\|^2 \tag{7.32}$$

The term $\mathbf{C}_{\theta_{\{k+1\}}}^{-1/2}$ can be rewritten using the Schur complement as

$$\mathbf{C}_{\theta_{\{k+1\}}}^{-1/2} = \begin{bmatrix} \mathbf{C}_{\theta_{\{k\}}}^{-1/2} & 0 \\ \mathbf{S}^{-1/2}\mathbf{C}_{\theta_{k+1}\theta_{\{k\}}}^{-1/2}\mathbf{C}_{\theta_{\{k\}}}^{-1/2} & \mathbf{S}^{-1/2} \end{bmatrix} \tag{7.33}$$

Substituting (7.33) into (7.32) leads to

$$\hat{\theta}_{\{k+1\}} = \arg\min_{\theta_{\{k+1\}}}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$
$$\left\|\left(\mathbf{J}_{\{k\}}^{\mathrm{H}}\mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1}\mathbf{J}_{\{k\}}\right)^{1/2}(\theta_{\{k\}} - \check{\theta}_{\{k\}})\right\|^2 +$$
$$\left\|\begin{bmatrix} \mathbf{C}_{\theta_{\{k\}}}^{-1/2} & 0 \\ \mathbf{S}^{-1/2}\mathbf{C}_{\theta_{k+1}\theta_{\{k\}}}^{-1/2}\mathbf{C}_{\theta_{\{k\}}}^{-1/2} & \mathbf{S}^{-1/2} \end{bmatrix}\left(\begin{bmatrix} \theta_{\{k\}} \\ \theta_{k+1} \end{bmatrix} - \begin{bmatrix} \bar{\theta}_{\{k\}} \\ \bar{\theta}_{k+1} \end{bmatrix}\right)\right\|^2 \tag{7.34}$$

$$\hat{\boldsymbol{\theta}}_{\{k+1\}} = \arg\min_{\boldsymbol{\theta}_{\{k+1\}}} (\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1})^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{k+1}}^{-1} (\mathbf{f}_{k+1}(\boldsymbol{\theta}_{k+1}) - \mathbf{x}_{k+1}) +$$

$$\left\| \begin{bmatrix} \left(\mathbf{J}_{\{k\}}^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} \mathbf{J}_{\{k\}} \right)^{1/2} & 0 \\ \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} & 0 \\ \mathbf{S}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} & \mathbf{S}^{-1/2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_{\{k\}} \\ \boldsymbol{\theta}_{k+1} \end{bmatrix} - \begin{bmatrix} \left(\mathbf{J}_{\{k\}}^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} \mathbf{J}_{\{k\}} \right)^{1/2} \breve{\boldsymbol{\theta}}_{\{k\}} \\ \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} \bar{\boldsymbol{\theta}}_{\{k\}} \\ \mathbf{S}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} \bar{\boldsymbol{\theta}}_{\{k\}} + \mathbf{S}^{-1/2} \bar{\boldsymbol{\theta}}_{k+1} \end{bmatrix} \right\|^2 \tag{7.35}$$

Lemma 7.2. *A term in a minimization problem over $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ of the form*

$$\left\| \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix} - \mathbf{r} \right\|^2 \tag{7.36}$$

can also be written as

$$\left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1 \right)^{\mathrm{T}} \mathbf{A}_1^{\mathrm{T}} \mathbf{A}_1 \left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_1 \right) + \left(\boldsymbol{\theta}_2 - \tilde{\boldsymbol{\theta}}_2 \right)^{\mathrm{T}} \mathbf{A}_2^{\mathrm{T}} \mathbf{P} \mathbf{A}_2 \left(\boldsymbol{\theta}_2 - \tilde{\boldsymbol{\theta}}_2 \right) \tag{7.37}$$

where

$$\tilde{\boldsymbol{\theta}}_1 = \left(\mathbf{A}_1^{\mathrm{T}} \mathbf{A}_1 \right)^{-1} \mathbf{A}_1^{\mathrm{T}} \left(\mathbf{r} - \mathbf{A}_2 \boldsymbol{\theta}_2 \right) \tag{7.38}$$

$$\tilde{\boldsymbol{\theta}}_2 = \left(\mathbf{A}_2^{\mathrm{T}} \mathbf{P} \mathbf{A}_2 \right)^{-1} \mathbf{A}_2^{\mathrm{T}} \mathbf{P} \mathbf{r} \tag{7.39}$$

$$\mathbf{P} = \mathbf{I} - \mathbf{A}_1 \left(\mathbf{A}_1^{\mathrm{T}} \mathbf{A}_1 \right)^{-1} \mathbf{A}_1^{\mathrm{T}} \tag{7.40}$$

Proof. Expanding (7.36) results in

$$(\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2)^{\mathrm{T}} (\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2) + (\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2)^{\mathrm{T}} \mathbf{r} + \mathbf{r}^{\mathrm{T}} (\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2) +$$
$$\mathbf{r}^{\mathrm{T}} \mathbf{r} \tag{7.41}$$

Expanding (7.37) results in

$$(\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2)^{\mathrm{T}} (\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2) + (\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2)^{\mathrm{T}} \mathbf{r} + \mathbf{r}^{\mathrm{T}} (\mathbf{A}_1 \boldsymbol{\theta}_1 + \mathbf{A}_2 \boldsymbol{\theta}_2) +$$
$$\mathbf{r}^{\mathrm{T}} \mathbf{P} \mathbf{A}_2 \left(\mathbf{A}_2^{\mathrm{T}} \mathbf{P} \mathbf{A}_2 \right)^{1} \mathbf{A}_2^{\mathrm{T}} \mathbf{P} \mathbf{r} \tag{7.42}$$

The only difference between (7.41) and (7.42) is the last term, which is independent of $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ hence for the purpose of minimization over $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, (7.36) and (7.37) are equivalent.   □

Now using Lemma 7.2 and defining

$$\mathbf{A}_1 = \begin{bmatrix} \left(\mathbf{J}_{\{k\}}^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} \mathbf{J}_{\{k\}} \right)^{1/2} \\ \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} \\ \mathbf{S}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} \end{bmatrix}, \quad \mathbf{A}_2 = \begin{bmatrix} 0 \\ 0 \\ \mathbf{S}^{-1/2} \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} \left(\mathbf{J}_{\{k\}}^{\mathrm{H}} \mathbf{C}_{\mathbf{w}_{\{k\}}}^{-1} \mathbf{J}_{\{k\}} \right)^{1/2} \breve{\boldsymbol{\theta}}_{\{k\}} \\ \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} \bar{\boldsymbol{\theta}}_{\{k\}} \\ \mathbf{S}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{k+1}\boldsymbol{\theta}_{\{k\}}}^{-1/2} \mathbf{C}_{\boldsymbol{\theta}_{\{k\}}}^{-1/2} \bar{\boldsymbol{\theta}}_{\{k\}} + \mathbf{S}^{-1/2} \bar{\boldsymbol{\theta}}_{k+1} \end{bmatrix} \tag{7.43}$$

and

$$\tilde{\theta}_{\{k\}} = \left(\mathbf{A}_1^{\mathsf{T}}\mathbf{A}_1\right)^{-1}\mathbf{A}_1^{\mathsf{T}}\left(\mathbf{r} - \mathbf{A}_2\theta_{k+1}\right) \tag{7.44}$$

$$\tilde{\theta}_{k+1} = \left(\mathbf{A}_2^{\mathsf{T}}\mathbf{P}\mathbf{A}_2\right)^{-1}\mathbf{A}_2^{\mathsf{T}}\mathbf{P}\mathbf{r} \tag{7.45}$$

$$\mathbf{P} = \mathbf{I} - \mathbf{A}_1\left(\mathbf{A}_1^{\mathsf{T}}\mathbf{A}_1\right)^{-1}\mathbf{A}_1^{\mathsf{T}} \tag{7.46}$$

the minimization program (7.35) can be written as

$$\hat{\theta}_{\{k+1\}} = \arg\min_{\theta_{\{k+1\}}}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathsf{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$

$$\left(\theta_{\{k\}} - \tilde{\theta}_{\{k\}}\right)^{\mathsf{T}}\mathbf{A}_1^{\mathsf{T}}\mathbf{A}_1\left(\theta_{\{k\}} - \tilde{\theta}_{\{k\}}\right) +$$

$$\left(\theta_{k+1} - \tilde{\theta}_{k+1}\right)^{\mathsf{T}}\mathbf{A}_2^{\mathsf{T}}\mathbf{P}\mathbf{A}_2\left(\theta_{k+1} - \tilde{\theta}_{k+1}\right) \tag{7.47}$$

From the equation above $\theta_{\{k\}}$ can be solved in closed form as function of θ_{k+1} yielding $\hat{\theta}_{\{k\}} = \tilde{\theta}_{\{k\}}$. The remaining problem

$$\hat{\theta}_{k+1} = \arg\min_{\theta_{\{k+1\}}}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1})^{\mathsf{H}}\mathbf{C}_{\mathbf{w}_{k+1}}^{-1}(\mathbf{f}_{k+1}(\theta_{k+1}) - \mathbf{x}_{k+1}) +$$

$$\left(\theta_{k+1} - \tilde{\theta}_{k+1}\right)^{\mathsf{T}}\mathbf{A}_2^{\mathsf{T}}\mathbf{P}\mathbf{A}_2\left(\theta_{k+1} - \tilde{\theta}_{k+1}\right) \tag{7.48}$$

is again a problem of the size of a single time block, but now the linearized version of all the previous steps is taken into account.

7.5   Conclusions

In this chapter we have briefly presented a few extensions to the single layer stochastic model discussed in the previous chapters. The models presented here are far from complete. They serve more as an example of how various correlations between parameters can be described. More observations are needed to formulate an verify such models. These observations will become available once LOFAR is fully operational. For all these models the MAP estimator retains the same structure of a non-linear least squares optimization problem. These solvers are already (being) implemented in the LOFAR software and hence the MAP estimator for more extended stochastic models can be readily implemented in the LOFAR software.

Chapter 8

Conclusions

Analysis of the LOFAR calibration problem, where the complex gain can vary rapidly over time, frequency and depending on direction, shows that there is no unique solution unless the gain is constrained to some low order model. For a polynomial model it has been shown that there is a unique solution and that the "peeling" method attains the Cramer Rao Lower Bound after a few iterations.

A derivation of a determinstic model for the ionosphere from first principles is infeasible. A phenomenological model inevitably involves some adhoc choices of the functions used to describe the ionopsphere and the model order.

However, a stochastic description of the ionosphere can be derived from first principles following the theory of turbulent flow by Kolmogorov. Furthermore the statistical properties of ionospheric fluctuations can be estimated from observations. Observational data shows that a powerlaw for the spectrum of ionospheric fluctuations is a reasonable model although the slope can deviate from the value predicted by Kolmogorov.

The optimal estimator in the least squares sense is the Bayesian Minimum Mean Square Error estimator. The Maximum A Posteriori (MAP) estimator is computationally feasible and achieves near optimal performance. Using the usual assumption of Gaussian process both for the ionospheric fluctuations and the noise the MAP estimator leads to a Least Squares (LS) problem. This problem can also be solved sequentially using the peeling method.

The combined MAP/Peeling method has been applied to VLA 74 MHz data and showed significant improvement over existing calibration methods. Other advantages include 1) the othogonality of the base vectors, which leads to less problems with convergence of the least squares solver and 2) the good behaviour of the basis functions at the edges unlike polynomials which results in proper extrapolation towards the edges of the field of view.

Possible extensions of this method include an extensions to a third spatial dimension and the inclusion of the time domain.

Bibliography

[1] *CRAF [Committee on Radio Astronomy Frequencies] Handbook for Radio Astronomy*. 2005.

[2] T.J. Cornwell and P.N.Wilkinson. A New Method for Making Maps with Unstable Radio Interferometers. *Mon. Not. R. Astron. Soc.*, 196:1067–1086, 1981.

[3] F. R. Schwab. Adaptive calibration of radio interferometer data. In W. T. Rhodes, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 231 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 18–25, January 1980.

[4] A.R. Thompson, J.M. Moran, and G.W. Swenson Jr. *Interferometry and Synthesis in Radio Astronomy, Second Edition*. Wiley-Interscience, New York, 2001.

[5] R.A. Perley, F. Schwab, and A.H. Bridle, editors. *Synthesis imaging in radio astronomy*. Astronomical society of the pacific, 1989.

[6] G.B. Taylor, C.L. Carilli, and R.A. Perley, editors. *Synthesis imaging in radio astronomy II : A collection of lectures from the sixth NRAO/NMIMT synthesis imaging summer school*. Astronomical Society of the Pacific, 1999.

[7] A.J. van der Veen, A. Leshem, and A.J. Boonstra. Array signal processing for radio astronomy. In P.J. Hall, editor, *The Square Kilometre Array: An Engineering Perspective*, pages 231–249. Springer, Dordrecht, 2005. URL `ftp://cas.et.tudelft.nl/pub/allejan-docs/aj04ska.pdf`. ISBN 1-4020-3797-x. Reprinted from Experimental Astronomy, 17(1-3),2004.

[8] G. B. Rybicki and A. P. Lightman. *Radiative processes in astrophysics*. New York, Wiley-Interscience, 1979. 393 p., 1979.

[9] C. J. Lonsdale. Configuration Considerations for Low Frequency Arrays. In N. Kassim, M. Perez, W. Junor, and P. Henning, editors, *Astronomical Society of the Pacific Conference Series*, volume 345 of *Astronomical Society of the Pacific Conference Series*, pages 399–+, December 2005.

[10] J.E. Noordam and A.G. de Bruyn. High dynamic range mapping of radio sources, with application to 3C84. *Nature*, 299:597–600, 1982.

[11] T. J. Pearson and A. C. S. Readhead. Image Formation by Self-Calibration in Radio Astronomy. *Annual Review of Astron and Astrophys*, 22:97–130, 1984. doi: 10.1146/annurev.aa.22.090184.000525.

[12] M.H. Wieringa. An Investigation of the Telescope Based Calibration Methods 'Redundancy' and 'Self-Cal'. *Experimental Astronomy*, 2:203–225, 1992.

[13] A.-J. Boonstra and A.-J. van der Veen. Gain Calibration Methods for Radio Telescope Arrays. *Signal Processing, IEEE Transactions on*, 51(1):25–38, January 2003.

[14] J.A. Hogbom. Aperture Synthesis with non-regular distribution of interferometer baselines. *Astronomy and Astrophysics, Supplement*, 15:417–426, 1974.

[15] U.J. Schwarz. Mathematical-statistical description of the iterative beam removing technique (method CLEAN). *Astronomy and Astrophysics*, 65:345–356, 1978.

[16] B. Friedlander and A.J. Weiss. Direction finding in the presence of mutual coupling. *IEEE Tr. Anten. Propag.*, 39(3):273–284, March 1991.

[17] B.C. Ng and C.M.S. See. Sensor-array calibration using a maximum-likelihood approach. *IEEE Tr. Anten. Prop.*, 44(6):827–835, June 1996.

[18] A. Flieller, A. Ferreol, P. Larzabal, and H. Clergeot. Robust bearing estimation in the presence of direction-dependent modelling errors: identifiability and treatment. In *IEEE ICASSP*, pages 1884–1887, April 1995.

[19] G.C. Brown, J.H. McClellan, and E.J. Holder. A homotopy continuation approach for self-calibration of arrays with general phase perturbations. In *IEEE Sixth SP Workshop on Stat. Signal Array Proc.*, pages 263–266, October 1992.

[20] A.J. Weiss and B. Friedlander. Array processing using joint diagonalization. *Signal Processing*, 50(3):205–222, May 1996.

[21] S.J. Wijnholds and A.-J. Boonstra. A multisource calibration method for phased array radio telescopes. pages 200–204, Boston (MA), July 2006. doi: 10.1109/SAM.2006.1706121.

[22] D.R. Fuhrmann. Estimation of sensor gain and phase. *IEEE Tr. Signal Proc.*, 42(1):77–87, January 1994.

[23] S. van der Tol, B.D. Jeffs, and A.-J.. van der Veen. Self-Calibration for the LOFAR Radio Astronomical Array. *Signal Processing, IEEE Transactions on*, 55(9):4497–4510, Sept. 2007. ISSN 1053-587X. doi: 10.1109/TSP.2007.896243.

[24] B. Jeffs, S. van der Tol, and A.J. van der Veen. Direction dependent self calibration of large distributed sensor arrays. In *IEEE ICASSP*, Toulouse (FR), May 2006.

[25] S. van der Tol, B. Jeffs, and A.J. van der Veen. Calibration of a large distributed low frequency radio astronomical array (LOFAR). In *EUSIPCO*, Antalya (T), sep 2005. Eurasip.

[26] J. E. Noordam. LOFAR calibration challenges. In *Proceedings of the SPIE, Vol. 5489, pp. 817-825 (2004).*, pages 817–825, October 2004.

[27] J. P. Hamaker, J. D. Bregman, and R. J. Sault. Understanding radio polarimetry. I. Mathematical foundations. *Astronomy and Astrophysics, Supplement*, 117:137–147, May 1996.

[28] J.E. Noordam. The MEASUREMENT EQUATION of a generic radio telescope, AIPS++ Implementation Note nr 185. Technical report, ASTRON, 1996.

[29] S.M. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice Hall, Upper Saddle River, New Jersy, 1993.

[30] S. J. Wijnholds, J. D. Bregman, and A. J. Boonstra. Sky Noise Limited Snapshot Imaging in the Presence of RFI with Lofar's Initial Test Station. *Experimental Astronomy*, 17:35–42, June 2004. doi: 10.1007/s10686-005-4949-9.

[31] A.S. Bennett. The revised 3C catalog of radio sources. *Memoires Royal Astron. Soc.*, 68:163–172, 1962.

[32] V. I. Tatarski. *Wave propagation in a turbulent medium; transl. from the Russian by R.A. Silverman.* New York, McGraw-Hill, 1961.

[33] F. Roddier. The Effects of Atmospheric Turbulence in Optical Astronomy. *Progress in Optics*, 19:281–376, 1981.

[34] A. Hewish. The Diffraction of Galactic Radio Waves as a Method of Investigating the Irregular Structure of the Ionosphere. *Proc. Roy. Soc. London A*, 214:494–514, 1952.

[35] B. Y. Mills and A. B. Thomas. Observations of the sources of radio-frequency radiation in the Constellation of Cygnus. *Australian Journal of Scientific Research A Physical Sciences*, 4:158–171, June 1951.

[36] D. G. Yerg. A Tentative Evaluation of Kinematic Viscosity for Ionospheric Regions. *Journal of Geophysics Research*, 57:217–+, June 1952.

[37] A. M. Prokhorov, F. V. Bunkin, K. S. Gochelashvili, and V. I. Shishov. Laser irradiance propagation in turbulent media. *IEEE Proceedings*, 63:790–811, May 1975.

[38] R. P. Linfield, M. M. Colavita, and B. F. Lane. Atmospheric Turbulence Measurements with the Palomar Testbed Interferometer. *Astrophysical Journal*, 554: 505–513, June 2001. doi: 10.1086/321372.

[39] K. C. Yeh and C.-H. Liu. Radio wave scintillations in the ionosphere. *IEEE Proceedings*, 70:324–360, 1982.

[40] P.F.J. van Velthoven. *Medium Scale Irregularities in the Ionospheric Electron Content*. PhD thesis, University of Eindhoven, Eindhoven, The Netherlands, 1990.

[41] A. S. Cohen, W. M. Lane, W. D. Cotton, N. E. Kassim, T. J. W. Lazio, R. A. Perley, J. J. Condon, and W. C. Erickson. The VLA Low-Frequency Sky Survey. *Astronomical Journal*, 134:1245–1262, September 2007. doi: 10.1086/520719.

[42] 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 J. M. Oschmann, Jr., editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 5489 of *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, pages 180–189, October 2004. doi: 10.1117/12.551298.

[43] A. S. Cohen and H. J. A. Röttgering. Probing Fine-Scale Ionospheric Structure with the Very Large Array Radio Telescope. 2008. to be submitted.

[44] D. L. Fried. Statistics of a Geometric Representation of Wavefront Distortion. *Journal of the Optical Society of America (1917-1983)*, 55:1427–1435, 1965.

[45] D. L. Fried. Differential Angle of Arrival: Theory, Evaluation and Measurement Feasibility. *Radio Science*, 10:71–76, 1975.

[46] D. L. Fried. Varieties of isoplanatism. In J. C. Wyant, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 75 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 20–29, 1976.

[47] K. A. Winick and D. V. Marquis. Stellar scintillation technique for the measurement of tilt anisoplanatism. *Journal of the Optical Society of America A*, 5:1929–1936, 1988.

[48] B. Ottersten, P. Stoica, and R. Roy. Covariance matching estimation techniques for array signal processing applications. *Digital Signal Processing*, 8:185–210, july 1998.

[49] E. A. Benediktov, V. D. Vyakhirev, N. P. Goncharov, L. V. Grishkevich, and V. A. Ivanov. Electron density variations in the ionospheric D layer. *Radiophysics and Quantum Electronics*, 21:238–240, March 1978. doi: 10.1007/BF01031684.

[50] I. Nagano and T. Okada. Electron Density Profiles in the Ionospheric D-Region Estimated from MF Radio Wave Absorption. *Advances in Space Research*, 25: 33–42, 2000. doi: 10.1016/S0273-1177(99)00894-7.

[51] S. van der Tol and A.-J. van der Veen. Ionospheric Calibration for the LOFAR Radio Telescope. volume 2, pages 1–4, Iasi, Romania, July 2007. doi: 10.1109/ISSCS.2007.4292761.

[52] C. J. Lonsdale, R. J. Cappello, J. E. Salah, J. N. Hewitt, M. F. Morales, L. J. Greenhill, R. Webster and D. Barnes. The Mileura Widefiels Array. In *Proceedings of the XXVIIIth General Assembly of the International Union of Radio Science (URSI GA)*, New Delhi, India, October 23-29 2005.

[53] J. B. Peterson, U. L. Pen and X. P. Wu. The Primeval Structure Telescope: Goals and Status. In *Proceedings of the XXVIIIth General Assembly of the International Union of Radio Science (URSI GA)*, New Delhi, India, October 23-29 2005.

[54] J. D. Bregman. LOFAR Approaching the Critical Design Review. In *Proceedings of the XXVIIIth General Assembly of the International Union of Radio Science (URSI GA)*, New Delhi, India, October 23-29 2005.

[55] W. D. Cotton and J.J. Condon. Calibration and Imaging of 74 MHz data from the Very large Array. In *URSI General Assembly*, Maastricht (NL), August 2002.

[56] Georges Matheron. Principles of geostatistics. *Economic Geology*, 58(8): 1246–1266, 1963. doi: 10.2113/gsecongeo.58.8.1246. URL `http://econgeol.geoscienceworld.org/cgi/content/abstract/58/8/1246`.

[57] A. R. Jacobson and W. C. Erickson. Wavenumber-resolved observations of ionospheric waves using the Very Large Array radiotelescope. *Planetary Space Science*, 40:447–455, April 1992. doi: 10.1016/0032-0633(92)90163-I.

[58] N. E. Kassim, T. J. W. Lazio, W. C. Erickson, R. A. Perley, W. D. Cotton, E. W. Greisen, A. S. Cohen, B. Hicks, H. R. Schmitt, and D. Katz. The 74 MHz System on the Very Large Array. *Astrophysical Journal, Supplement*, 172:686–719, October 2007. doi: 10.1086/519022.

[59] G. Swarup. Giant metrewave radio telescope (GMRT). In T. J. Cornwell and R. A. Perley, editors, *IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*, volume 19 of *Astronomical Society of the Pacific Conference Series*, pages 376–380, 1991.

[60] K. Davies. *Ionospheric Radio*. Institution of Engineering and Technology, Stevenage, UK, 1990.

[61] N. A. B. Gizani, A. Cohen, and N. E. Kassim. First results of the 74-MHz Very Large Array-Pie Town link. Hercules A at low frequencies. *Monthly Notices of the RAS*, 358:1061–1068, April 2005. doi: 10.1111/j.1365-2966.2005.08849.x.

[62] H.J. van Langevelde and R.M. Campbell. Evaluation of Ionospheric Correction Method for the European VLBI Network. In *URSI General Assembly*, New Delhi (India), October 2005. http://www.ursi.org/Proceedings/ProcGA05/pdf/J03-P.8(01310).pdf.

[63] J. E. Noordam. The MIM principle – version 2.0. Technical report, 2008. in preparation, available on request through *noordam@astron.nl*.

[64] F. R. Schwab. Relaxing the isoplanatism assumption in self-calibration; applications to low-frequency radio interferometry. *Astronomical Journal*, 89:1076–1081, July 1984. doi: 10.1086/113605.

[65] C. R. Subrahmanya. Low frequency imaging and the non-isoplanatic atmosphere. In T. J. Cornwell and R. A. Perley, editors, *IAU Colloq. 131: Radio Interferometry. Theory, Techniques, and Applications*, volume 19 of *Astronomical Society of the Pacific Conference Series*, pages 218–222, 1991.

[66] A. S. Cohen, H. J. A. Röttgering, N. E. Kassim, W. D. Cotton, R. A. Perley, R. Wilman, P. Best, M. Pierre, M. Birkinshaw, M. Bremer, and A. Zanichelli. The Low-Frequency Radio Counterpart of the XMM Large-Scale Structure Survey. *Astrophysical Journal*, 591:640–661, July 2003. doi: 10.1086/375397.

[67] A. S. Cohen, H. J. A. Röttgering, M. J. Jarvis, N. E. Kassim, and T. J. W. Lazio. A Deep, High-Resolution Survey at 74 MHz. *Astrophysical Journal, Supplement*, 150:417–430, February 2004. doi: 10.1086/380783.

[68] H. J. A. Rottgering, R. Braun, P. D. Barthel, M. P. van Haarlem, G. K. Miley, R. Morganti, I. Snellen, H. Falcke, A. G. de Bruyn, R. B. Stappers, W. H. W. M. Boland, H. R. Butcher, E. J. de Geus, L. Koopmans, R. Fender, J. Kuijpers, R. T. Schilizzi, C. Vogt, R. A. M. J. Wijers, M. Wise, W. N. Brouw, J. P. Hamaker, J. E. Noordam, T. Oosterloo, L. Bahren, M. A. Brentjens, S. J. Wijnholds, J. D. Bregman, W. A. van Cappellen, A. W. Gunst, G. W. Kant, J. Reitsma, K. van der Schaaf, and C. M. de Vos. LOFAR - Opening up a new window on the Universe. *ArXiv Astrophysics e-prints*, October 2006.

[69] T.J.W. Lazio, N.E. Kassim, and R.A. Perley. Low-Frequency Data Reduction at the VLA: A Tutorial for New Users, version 1.13. Technical report, 2005. URL `http://lwa.nrl.navy.mil/tutorial/tutorial.html`.

[70] A.H. Bridle and E.W. Greisen. The NRAO AIPS Project – A Summary, AIPS Memo 87. Technical report, NRAO, April 1994. URL `ftp://ftp.aoc.nrao.edu/pub/software/aips/TEXT/PUBL/AIPSMEMO87.PS`.

[71] M. Kettenis, H. J. van Langevelde, C. Reynolds, and B. Cotton. ParselTongue: AIPS Talking Python. In C. Gabriel, C. Arviset, D. Ponz, and S. Enrique, editors, *Astronomical Data Analysis Software and Systems XV*, volume 351 of *Astronomical Society of the Pacific Conference Series*, pages 497–+, July 2006.

[72] R. A. Perley. Wide Field Imaging II: Imaging with Non-Coplanar Arrays. In R. A. Perley, F. R. Schwab, and A. H. Bridle, editors, *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*, pages 259–+, 1989.

[73] T. J. Cornwell and R. A. Perley. Radio-interferometric imaging of very large fields - The problem of non-coplanar arrays. *Astronomy and Astrophysics*, 261:353–364, July 1992.

[74] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and Flannery B.P. *Numerical Recipes in C, Second Edition*. Cambridge University Press, Cambridge, UK, 1992.

[75] Craig B. Markwardt. Non-linear Least Squares Fitting in IDL with MPFIT. 2008.

[76] F.R. Schwab. VLA Scientific Memo 136. Technical report, 1981. URL `http://www.vla.nrao.edu/memos/sci/136.pdf`.

[77] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, and J. J. Broderick. The NRAO VLA Sky Survey. In D. R. Crabtree, R. J. Hanisch, and J. Barnes, editors, *Astronomical Data Analysis Software and Systems III*, volume 61 of *Astronomical Society of the Pacific Conference Series*, pages 155–+, 1994.

[78] J. J. Condon, W. D. Cotton, E. W. Greisen, Q. F. Yin, R. A. Perley, G. B. Taylor, and J. J. Broderick. The NRAO VLA Sky Survey. *Astronomical Journal*, 115:1693–1716, May 1998. doi: 10.1086/300337.

[79] R. B. Rengelink, Y. Tang, A. G. de Bruyn, G. K. Miley, M. N. Bremer, H. J. A. Roettgering, and M. A. R. Bremer. The Westerbork Northern Sky Survey (WENSS), I. A 570 square degree Mini-Survey around the North Ecliptic Pole. *Astronomy and Astrophysics, Supplement*, 124:259–280, August 1997. doi: 10.1051/aas:1997358.

[80] R. A. Perley. High Dynamic Range Imaging. In R. A. Perley, F. R. Schwab, and A. H. Bridle, editors, *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*, pages 287–+, 1989.

[81] D. S. Briggs and T. J. Cornwell. An Alternative Interpretation for the Physical Basis of CLEAN. In D. M. Worrall, C. Biemesderfer, and J. Barnes, editors, *Astronomical Data Analysis Software and Systems I*, volume 25 of *Astronomical Society of the Pacific Conference Series*, pages 170–+, 1992.

[82] D.S. Briggs. *High Fidelity Deconvolution of Moderately Resolved Sources*. PhD thesis, New Mexico Institute of Mining Technology, Socorro, New Mexico, USA, 1995.

[83] M. A. Voronkov and M. H. Wieringa. The Cotton-Schwab Clean At Ultra-High Dynamic Range. *Experimental Astronomy*, 18:13–29, April 2004. doi: 10.1007/s10686-005-9000-7.

[84] W.D. Cotton and J.M.Uson. EVLA Memo 114. Technical report, NRAO, 2007. URL http://www.aoc.nrao.edu/evla/geninfo/memoseries/evlamemo114.pdf.

[85] W. D. Cotton. Special Problems in Imaging. In R. A. Perley, F. R. Schwab, and A. H. Bridle, editors, *Synthesis Imaging in Radio Astronomy*, volume 6 of *Astronomical Society of the Pacific Conference Series*, pages 233–+, 1989.

[86] T. Cornwell, R. Braun, and D. S. Briggs. Deconvolution. In G. B. Taylor, C. L. Carilli, and R. A. Perley, editors, *Synthesis Imaging in Radio Astronomy II*, volume 180 of *Astronomical Society of the Pacific Conference Series*, pages 151–+, 1999.

[87] B. G. Clark. An efficient implementation of the algorithm 'CLEAN'. *Astronomy and Astrophysics*, 89:377–+, September 1980.

[88] S. Bhatnagar, T. J. Cornwell, K. Golap, and J. M. Uson. Correcting direction-dependent gains in the deconvolution of radio interferometric images. *Astronomy and Astrophysics*, 487:419–429, August 2008. doi: 10.1051/0004-6361:20079284.

[89] DoD World Geodetic System 1984. Technical Report TR8350.2, NIMA, 1984.

[90] N.R. Mohan. ANAAMIKA manual – version 2.1. Technical report, 2008. URL http://www.strw.leidenuniv.nl/$\sim$mohan/anaamika_manual.pdf.

[91] J. J. Condon. Errors in Elliptical Gaussian FITS. *Publications of the ASP*, 109:166–172, February 1997. doi: 10.1086/133871.

[92] R. H. Becker, R. L. White, and D. J. Helfand. The FIRST Survey: Faint Images of the Radio Sky at Twenty Centimeters. *Astrophysical Journal*, 450:559–+, September 1995. doi: 10.1086/176166.

[93] M. J. Jarvis, S. Rawlings, C. J. Willott, K. M. Blundell, S. Eales, and M. Lacy. On the redshift cut-off for steep-spectrum radio sources. *Monthly Notices of the RAS*, 327:907–917, November 2001. doi: 10.1046/j.1365-8711.2001.04778.x.

[94] T. A. Enßlin and H. Röttgering. The radio luminosity function of cluster radio halos. *Astronomy and Astrophysics*, 396:83–89, December 2002. doi: 10.1051/0004-6361:20021382.

[95] R. Cassano, G. Brunetti, and G. Setti. Statistics of giant radio haloes from electron reacceleration models. *Monthly Notices of the RAS*, 369:1577–1595, July 2006. doi: 10.1111/j.1365-2966.2006.10423.x.

[96] JOHN WISHART. THE GENERALISED PRODUCT MOMENT DISTRIBUTION IN SAMPLES FROM A NORMAL MULTIVARIATE POPULATION. *Biometrika*, 20A(1-2):32–52, 1928. doi: 10.1093/biomet/20A.1-2.32. URL http://biomet.oxfordjournals.org.

Summary

Radio astronomical observations at low frequencies (< 250 MHz), can be severely distorted by fluctuations in electron density in the ionosphere. The free electrons cause a phase change of electromagnetic waves traveling through the ionosphere. This effect increases for lower frequencies. For this reason observations at low frequencies have been limited to short baselines and hence poor angular resolution. Most radio astronomical observations today are done at higher frequencies.

The lower frequency bands however contain signals that are of great scientific value. Due to the expansion of the universe signals from distant objects are redshifted, i.e. shifted to lower frequencies. The more distant an observed object is the further we look back in time. An important period in the history of the universe is the "Epoch of Reionization" (EoR). A few hundred thousand years after the Big Bang the universe has cooled enough to allow the formation of neutral hydrogren and helium. When the universe was a few hundred million years old the EoR started and the almost completely neutral gas was ionized again. Probably the only method to trace the neutral gas in this period is through the 1420 MHz spectral line of neutral hydrogen. For the EoR this line is redshifted to somewhere probably 100 and 200 MHz.

Recently it has in principle become possible to observe with high resolution at low frequencies because the ever increasing computing power of digital data processing devices has made it possible to correct for the effect of the ionosphere. Determining the necessary corrections to the effect of the ionosphere is calles ionospheric calibration. The reason that ionospheric calibration is difficult is that the gain is direction dependent and rapidly varying over time. This greatly increases the number of degrees of freedom which causes two problems. First, the estimation of a large number of parameters is computationally costly. Second, the more parameters need to be estimated, the larger the estimation error will be. Without further constraints the signal to noise ratio of the calibrator sources is too low to accurately estimate the free parameters.

The first problem can in principle be tackled by a brute force approach by simply increasing the data processing capacity. In practice an efficient algorithm is needed. The second problem is more fundamental in nature. A good model of the ionosphere, including as much prior knowledge about the ionosphere as possible, is needed to reduce the number of degrees of freedom.

The first problem is addressed in this thesis by an analysis of a proposed calibration method called "Peeling". This is a calibration technique whereby the Least Squares (LS) optimization problem is sequentially solved for different calibrator sources. This can be computationally more efficient than joint estimation. Our analysis by simulation of a realistic target field finds that "Peeling" reaches the theoretically optimal result in a few iterations.

The second problem is addressed by proposing a stochastic ionospheric model based on a single layer of Kolmogorov turbulence. The stochastic model consists of a parametric description of the spatial power spectral density of the ionospheric electron density fluctuations. This model is verified by GPS observations and low frequency observations from the Very Large Array (VLA).

An optimal estimator for this model is the Bayesian Minimum Mean Square Error (MMSE) estimator. This estimator is impractical due to the necessary numerical integration of high dimensional integrals. The Maximum a Posteriori (MAP) estimator is an approximation of the MMSE estimator which leads to a Least Squares (LS) problem that can be solved efficiently by standard techniques. Simulations show that the MAP estimator based on the power law model performs better than estimation based on a Zernike polynomial model for the ionosphere.

The MAP estimator has been incorporated into the software package SPAM (Source Peeling and Atmospheric Modeling). SPAM has been used on three test cases, a simulated visibility data set and two selected 74 MHz VLA data sets. This resulted in significant improvements in image background noise (5–75 percent reduction) and source peak fluxes (up to 25 percent increase) as compared to the existing self-calibration and field-based calibration methods. The improved image quality indicates a significant improvement in ionospheric phase calibration accuracy.

For this particular single layer ionospheric model the results are encouraging. It is indicated how the MAP estimator can be applied to possible extensions of the model including the addition of a third spatial dimension and the time dimension.

Samenvatting

Radioastronomische waarnemingen op lage frequenties (< 250 MHz) kunnen ernstig verstoord worden door fluctuaties in de elektrondichtheid in de ionosfeer. De vrije elektronen veroorzaken een fase verdraaing van elektromagnetische golven die door de ionosfeer reizen. Des te lager de frequentie van de golf des te groter dit effect. Om deze reden zijn waarneming op lage frequenties tot nu toe beperkt gebleven tot korte basislijnen en dus een lage resolutie.

De laag frequente signalen bevatten echter informatie van grote wetenschappelijke waarde. Door de expansie van het heelal worden de signalen van ver weg gelegen objecten roodverschoven, d.w.z. verschoven naar lagere frequenties. Des te verder een object van ons is verwijderd, des te verder kijken we terug in de tijd en des te meer zijn de signalen in frequentie verschoven. Een belangrijke periode in de geschiedenis van het heelal is de zogenaamde "Epoch of Reionization" (EoR). Enkele honderdduizenden jaren na de Big Bang was het heelal ver genoeg afgekoeld zodat neutraal waterstof- en heliumgas kon worden gevormd. Toen het heelal enkele honderden miljoenen jaren oud was begon de EoR werd het vrijwel geheel neutrale gas in het universum opnieuw geioniseerd. Over hoe en wanneer dit precies gebeurde is nog weinig bekend. Waarschijnlijk de enige manier om het neutrale gas gedurende deze periode te traceren is via de 1420 MHz spectrale lijn van neutraal waterstof. Voor de EoR bevind deze lijn zich door de roodverschuiving ergens tussen de 100 en 200 MHz.

Recent is het in principe mogelijk geworden om op deze frequenties met hoge resolutie waar te nemen omdat het dankzij de toegenomen verwerkingscapaciteit van digitale hardware mogelijk is geworden te corrigeren voor het effect van de ionosfeer. Het bepalen van de benodigde correcties wordt ionosferische calibratie genoemd. De reden dat de ionosferische calibratie zo moeilijk is, is dat de fase verdraaing ten gevolge van de ionosfeer sterk richtingsafhankelijk is en sterk varieert over tijd. Dit verhoogt het aantal vrijheidsgraden van het calibratieprobleem aanzienlijk. Ten eerste nemen hierdoor de kosten in rekenkracht sterk toe en ten tweede neemt de nauwkeurigheid van de schatting af. Zonder aannames over de ionosfeer is calibratie in veel gevallen onmogelijk.

Het eerste probleem kan in principe opgelost worden door voldoende rekenkracht in te zetten. In de praktijk zijn efficiente algoritmes nodig om tot een acceptabele rekentijd te komen. Het tweede probleem is meer fundamenteel van aard. Zonder een goed model dat zoveel mogelijk a priori kennis van de ionosfeer bevat is calibratie niet mogelijk zelfs als de rekencapaciteit ongelimiteerd is.

Een eerder voorgesteld potentieel efficient algoritme genaamd "Peeling" wordt in deze thesis geanalyseerd. Dit algoritme lost het kleinste kwadraten probleem sequentieel per calibrator bron op. Onze analyse met behulp van een simulatie van een realitisch bronnen veld laat zien dat dit algoritme binnen enkele iteraties convergeert en de theoretische ondergrens voor de schattingsfout behaalt.

Voor het modeleren van de ionosfeer is een stochastisch model gekozen. De ionosferische fase fluctuaties worden beschreven door een enkele laag van Kolmogorov turbulentie. Het stochastisch model bestaat uit een parametrische beschrijving van het spectrum van de fluctuaties van de elektron dichtheid in de ionosfeer. Het model wordt gever-

ifieerd met behulp van data van GPS satellieten en observaties met het 74 MHz systeem van de Very Large Array (VLA).

Een optimale schatter voor een stochastisch data model is de Bayesiaanse kleinste kwadraten schatter. In de praktijk is deze schatter moeilijk de realiseren omdat numerieke integratie noodzakelijk is. Een benadering is the Maximum A Posteriori (MAP) schatter. Deze schatter leidt weer tot een kleinste kwadraten probleem voor een Gaussisch data model. Simulaties tonen aan dat voor een dergelijk model de MAP schatter een kleinere schattingsfout oplevert dan voor een kleinste kwadraten fit aan een deterministisch model gebaseerd op Zernike polynomen.

Het calibratie pakket Source Peeling and Atmospheric Modeling (SPAM) combineert "Peeling" met het stochastisch model. SPAM is toegepast op drie scenario's, een gesimuleerde dataset en twee observaties op 74 MHz van de VLA. Het resultaat toont een significante reductie (5–75 procent) van de achtergrondruis in de kaart en een hogere peak flux van de radiobronnen (toename tot 25 procent) vergeleken met standaard zelf-calibratie en Field Based Calibration. Het is niet mogelijk om de nauwkeurigheid van fase oplossingen direct te bepalen, omdat de echte ionosferische fase onbekend is. De verbeterde kwaliteit van de kaart is een indicate voor een significante verbetering van de nauwkeurigheid van de ionosferische calibratie.

De resultaten voor dit ionosferisch model en de MAP schatter zijn bemoedigend. Daarom is verder nog aangegeven hoe de MAP schatter toegepast kan worden op mogelijke uitbreidingen van het model zoals het toevoegen van een derde spatiele dimensie en de tijd dimensie.

About the author

Sebastiaan van der Tol was born in The Netherlands in 1977. He received the M.Sc. degree in Electrical Engineering from Delft University of Technology, The Netherlands, in 2004. From January 2004 to September 2008 he has been a research assistent with the same institute, where he is pursuing a Ph.D. degree in Electrical Engineering. Since May 2009 he is a research associate with Leiden Observatory, Leiden University, The Netherlands. His current research interests include array signal processing and interference mitigation techniques for large phased array radio telescopes.