Ionospheric Calibration for the LOFAR Radio Telescope

Sebastiaan van der Tol and Alle-Jan van der Veen
TU Delft, Fac. EEMCS, 2628 CD Delft, The Netherlands
Email: S.vanderTol@tudelft.nl, A.J.vanderVeen@tudelft.nl

***Abstract*— 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-Loeve 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.**

I. 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) [1], the Primeval Structure Telescope (PaST) [2] and the Low Frequency Array (LOFAR) [3] which we consider in this paper. 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 [4], and as described in [5] 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 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. 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 paper 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 popular method of fitting Zernike polynomials [6].

[0]This work was supported by NWO-STW (The Netherlands) under grant number DTC.5893.

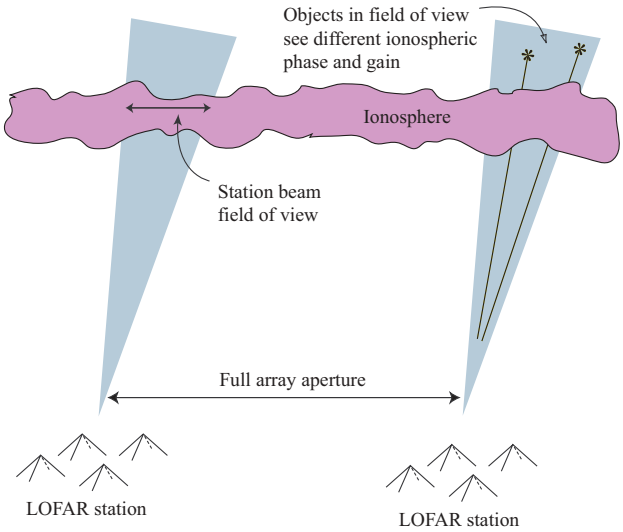

Fig. 1. LOFAR scenario: the ionosphere changes within the station beam and direction-dependent calibration is required. Also the ionosphere is different for different stations.

II. Data model and Problem Statement

A. 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

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

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

$$\mathbf{x}_k[n] = \mathbf{a}_k s_k[n] + \mathbf{w}_k[n]\,, \tag{2}$$

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

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

where

$$\boldsymbol{\phi}_k = C\boldsymbol{\tau}_k f_k^{-1} \tag{3}$$

is a vector with M entries representing the ionospheric phases at each station (a function of frequency), and $\boldsymbol{\tau}_k$ 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 $C = 8422\,\mathrm{rad/MHz/TECU}$ (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.

B. 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 [7]

$$D_{\mathrm{TEC}}(r) = \left(\frac{f_0}{C}\right)^2 \left(\frac{r}{s_0}\right)^\beta \tag{4}$$

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 $D_\phi(s_0, f_0) = 1$.

III. Calibration Algorithm

A. General Data Model

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

$$\mathbf{y} = \mathbf{f}(\boldsymbol{\theta}) + \mathbf{w} \tag{5}$$

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

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

where $\mathrm{vec}(\cdot)$ stacks the columns of a matrix into a vector, and $\otimes$ denotes the Kronecker product. The observation noise $\mathrm{vec}(\hat{\mathbf{R}}_k - \mathbf{R}_k)$ corresponds to the noise vector $\mathbf{w}$ in (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 $\boldsymbol{\tau}_k$, but since we cannot expect to estimate the bulk delay, we subtract the average value of $\boldsymbol{\tau}_k$ and define

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

The corresponding covariance matrix for $\boldsymbol{\theta}$ is, from (4),

$$\mathbf{C}_\theta = \frac{-1}{2}\left(\frac{f_0}{C}\right)^2 (\mathbf{I} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}) \left(\frac{\mathbf{D}}{s_0}\right)^{\odot\beta} (\mathbf{I} - \frac{1}{M}\mathbf{1}\mathbf{1}^{\mathrm{T}}) \tag{6}$$

where $\mathbf{D}$ 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 $\mathbf{C}_\theta$.

B. MMSE estimator

A desirable estimator is the estimator with the minimum Mean Squared Error, the MMSE, given by [8]

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

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.

C. MAP estimator

A good alternative is to use the Maximum A Posteriori (MAP) estimator [8],

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

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

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

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

so that

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max_{\boldsymbol{\theta}} p(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &= \arg\min_{\boldsymbol{\theta}} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{y}-\mathbf{f}(\boldsymbol{\theta}))\|^2 + \|\mathbf{C}_{\theta}^{-\frac{1}{2}}\boldsymbol{\theta}\|^2\,. \quad (7)\end{aligned}$$

D. Karhunen-Loeve transformation

In many cases the dimensionality of $\boldsymbol{\theta}$ is large. It may contain not only entries corresponding to sample points, but also entries corresponding to distances where we wish to know the phase or the TEC, e.g., on a uniformly spaced grid. Thus, it is necessary to interpolate the sample points, assuming a "smooth" behavior over space and taking into account the spatial correlation behavior given by the structure function, or equivalently the covariance matrix $\mathbf{C}_{\theta}$.

The full parameter vector $\boldsymbol{\theta}$ is "explained" in terms of fewer underlying parameters by posing

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

where $\mathbf{U}_{\theta}$ is a tall matrix whose columns are regarded as basis functions of the interpolation. 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

$$\mathbf{C}_{\theta} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^{\mathrm{H}} \approx \mathbf{U}_{\theta}\boldsymbol{\Lambda}_{\theta}\mathbf{U}_{\theta}^{\mathrm{H}} \quad (8)$$

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

Inserting this into the MAP estimator, we obtain

$$\hat{\mathbf{p}} = \arg\min_{\mathbf{p}} \|\mathbf{C}_{\mathbf{w}}^{-\frac{1}{2}}(\mathbf{y}-\mathbf{f}(\mathbf{U}_{\theta}\hat{\mathbf{p}}))\|^2 + \|\boldsymbol{\Lambda}_{\theta}^{-\frac{1}{2}}\hat{\mathbf{p}}\|^2\,. \quad (9)$$

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

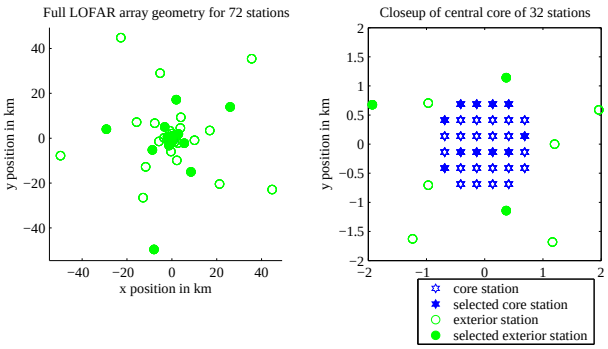

Fig. 2. Station configuration used for simulation. Only the selected stations were actually included in the simulation.

E. 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 $\boldsymbol{\theta}$, 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 (9)
3) Estimate β and s_0 using a non-linear solver as

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

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

IV. 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 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 (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 (2). The covariance estimates are then obtained from (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.

Based on the simulated data the TEC values are estimated using three different methods: Least Squares using Zernike polynomials, Least Squares using a Karhunen-Loeve basis, and the MAP using a Karhunen-Loeve basis. All methods are based on equation (9). When we omit the term $\|\boldsymbol{\Lambda}_\theta^{-\frac{1}{2}}\hat{\mathbf{p}}\|$, the method reduces to an ordinary least squares fit. The basis $\mathbf{U}_\theta$ consists either of Zernike polynomials or the Karhunen-Loeve basis computed from (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 $\boldsymbol{\tau}_k$ is estimable.

From the estimated $\boldsymbol{\tau}_k$, the ionospheric phases $\boldsymbol{\phi}_k$ are computed as in (3). The error measure is the r.m.s. phase error at the reference frequency $f_0 = 100$MHz.

In figure $3(a)$ 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 (8) improves the performance. The lowest error of the LS method using the Karhunen-Loeve 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 $3(b)$, 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 III-E.

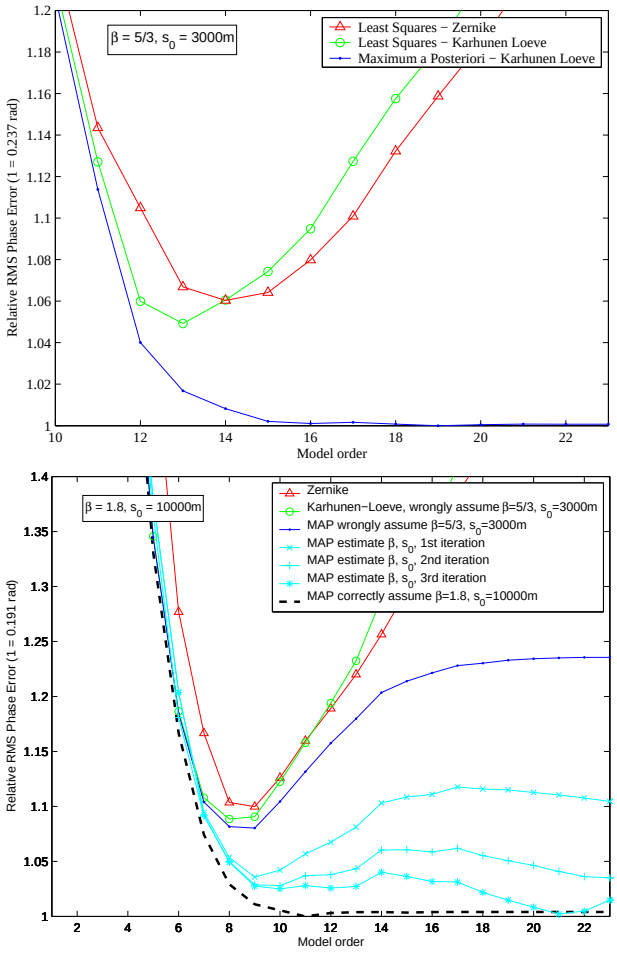

Fig. 3. Estimation performance as function of model order selection: (a) known model parameters, (b) estimated model parameters. The error is scaled relative to the minimum in each plot.

REFERENCES

[1] 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), Oct. 23-29 2005.

[2] 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), Oct. 23-29 2005.

[3] 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), Oct. 23-29 2005.

[4] J.E. Noordam, "LOFAR calibration challenges," in *Proc. SPIE*, vol. 5489, pp. 817–825, Oct. 2004.

[5] S. van der Tol, B.D. Jeffs, and A.J. van der Veen, "Self calibration for the LOFAR radio astronomical array," *To appear in IEEE Tr. Signal Processing*, 2007.

[6] W.D. Cotton and J.J. Condon, "Calibration and imaging of 74 mhz data from the very large array," in *XXVIIth General Assembly of the International Union of Radio Science, URSI 2002*, 2002.

[7] A.R. Thompson, J.M. Moran, and G.W. Swenson, eds., *Interferometry and Synthesis in Radioastronomy*. John Wiley and Sons, second ed., 2001.

[8] S.M. Kay, *Fundumentals of statistical signal processing: Estimation theory*. PTR, Prentice Hall, 1993.