

# CRB Analysis of the Impact of Unknown Receiver Noise on Phased Array Calibration

Sebastiaan van der Tol  
Delft Univ. of Technology  
Dept. Electrical Eng./DIMES  
Mekelweg 4  
2628 CD Delft  
The Netherlands  
Email: S.vanderTol@tudelft.nl

Stefan J. Wijnholds  
ASTRON  
R&D Department  
Oude Hoogeveensedijk 4  
7991 PD Dwingeloo  
The Netherlands  
Email: wijnholds@astron.nl

**Abstract**—In radio astronomy images are made of astronomical objects as they appear at radio frequencies using a technique called aperture synthesis. Signals from several antennas are correlated and integrated over time. The data collected over several hours are further processed to calibrate the instrument and to form an image or intensity map. The calibration and imaging algorithms do not use the autocorrelations because the receiver noise is unstable and hence considered unknown. In literature the Cramer Rao Bound for the calibration problem has been derived assuming that the autocorrelations are part of the available data. If the assumption is correct that the autocorrelations do not contain useful information when the receiver noise is unknown, than the CRB for the case that the autocorrelations are not part of the data will be the same. In this paper we will derive the CRB excluding the autocorrelations and show that it indeed does not matter whether the autocorrelations are included or not.

## 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]. At these frequencies the beamwidths of the receiving elements are wider than those of most existing radio telescopes, therefore there are always multiple bright sources present within the field of view. Existing calibration techniques based on a single source calibration can not be used. Furthermore the effects of the ionosphere are stronger at lower frequencies, which makes a direction dependent calibration necessary.

New algorithms are needed for calibration at low frequencies.

During our work on calibration algorithms for LOFAR we have noticed that the Cramer Rao Bound for the radioastronomical calibration problem has been derived assuming an estimate of the full covariance matrix,  $\hat{\mathbf{R}}$ , is available, while existing calibration algorithms discard the diagonal of  $\hat{\mathbf{R}}$  [4]. Discarding the diagonal makes sense because the noise powers are considered to be unknown: the data model for the covariance matrix is of the form  $\mathbf{R} = \mathbf{R}(\boldsymbol{\theta}) + \mathbf{D}$ , where  $\boldsymbol{\theta}$  are the unknown calibration parameters (the gains and phase offsets) and  $\mathbf{D}$  is a diagonal matrix containing the unknown noise powers. Existing algorithms discard the diagonal of  $\hat{\mathbf{R}}$  and only estimate  $\boldsymbol{\theta}$ . Of course, the lower bound for the case that the diagonal of  $\hat{\mathbf{R}}$  is available is also a lower bound for the case the diagonal has been discarded, because discarding data can only increase the variance. But the question remains whether the CRB is tighter (higher) if we take into account that the diagonal has been discarded. If so, the diagonal contains some useful information to estimate  $\boldsymbol{\theta}$  and there might exist an algorithm that exploits that. Our intuition was that that is not the case. In this paper we present a proof that the bounds are indeed the same.

Section II states the problem in mathematical terms. In section III we will motivate our intuition. Section IV gives the necessary background to formally solve the problem. Section V presents the results.

*Notation:* Vectors are denoted by bold lowercase,  $\mathbf{x}$ . Matrices by bold capitals  $\mathbf{X}$ . The transpose operator is denoted by  $^T$  and the conjugate transpose by  $^H$ . Entries of matrix  $\mathbf{X}$  are denoted by  $x_{i,j}$ . For entries of matrix expressions we use a combination brackets and indices  $[\mathbf{X}^{-1}]_{i,j}$ .

## II. DATA MODEL & PROBLEM STATEMENT

Assume we have an array of  $M$  elements. The sampled array output in complex base band form for the  $n^{\text{th}}$  sample is modeled as

$$\mathbf{x}[n] = \mathbf{x}_s[n] + \mathbf{n}[n]$$

where  $\mathbf{x}[n]$  is the  $M \times 1$  output vector,  $\mathbf{x}_s[n]$  is the signal from the astronomical sources and  $\mathbf{n}[n]$  is the noise. The statistical model for the signals is  $\mathbf{x}_s[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}))$  where  $\boldsymbol{\theta}$  are the unknown calibration parameters, and  $\mathbf{n}[n] \sim \mathcal{CN}(\mathbf{0}, \mathbf{D})$  where  $\mathbf{D}$  is a diagonal matrix of which the entries can be considered either known or unknown. The array output is correlated and averaged over time to form

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

The signal and the noise are assumed to be independent so

$$\mathbb{E}[\hat{\mathbf{R}}] = \mathbf{R}(\boldsymbol{\theta}) + \mathbf{D}.$$

This is a general model which fits more specific commonly used models [4]–[7].

The question we will try to answer is the following. Suppose the diagonal entry  $d_{i,i}$  is unknown, does knowledge of the corresponding data entry  $\hat{r}_{i,i}$  help us to estimate  $\boldsymbol{\theta}$  or can  $\hat{r}_{i,i}$  simply be discarded without influencing the maximum attainable performance? Algorithms as in [8] do not use the diagonal of  $\hat{\mathbf{R}}$ ; the noise powers  $\mathbf{D}$  are unknown, but not estimated. The algorithms in [4] do use the diagonal but they also estimate  $\mathbf{D}$ . The question is whether these two approaches are equivalent or not.

## III. INTUITIVE ANSWER BASED ON THE LEAST SQUARES ALGORITHM

An intuitive answer can be found when we try to solve this estimation problem with a least squares fit and assume that that will give us the optimal result.

The least squares fit equation, for known  $\{d_{i,i}\}, i = 1 \dots M$ , is given by

$$\begin{aligned} \boldsymbol{\theta} &= \arg \min_{\boldsymbol{\theta}} \|\hat{\mathbf{R}} - (\mathbf{R}(\boldsymbol{\theta}) + \mathbf{D})\|_F^2 \\ &= \arg \min_{\boldsymbol{\theta}} \sum_{j=1}^M \sum_{k=1}^M (\hat{r}_{j,k} - (r(\boldsymbol{\theta})_{j,k} + d_{j,k}))^2, \end{aligned}$$

or, for unknown  $\{d_{i,i}\}, i = 1 \dots M$ ,

$$\begin{aligned} \{\boldsymbol{\theta}, d_{i,i}\} &= \arg \min_{\boldsymbol{\theta}, d_{i,i}} \|\hat{\mathbf{R}} - (\mathbf{R}(\boldsymbol{\theta}) + \mathbf{D})\|_F^2 = \\ &= \arg \min_{\boldsymbol{\theta}, d_{i,i}} \sum_{j=1}^M \sum_{k=1}^M (\hat{r}_{j,k} - (r(\boldsymbol{\theta})_{j,k} + d_{j,k}))^2, \end{aligned}$$

First consider the case of unknown  $\{d_{i,i}\}$ . The unknown  $d_{i,i}$  appears only in one term of the sum so we are free to choose  $d_{i,i}$  (for each  $i$ ) such that

$$(\hat{r}_{j,k} - (r(\boldsymbol{\theta})_{i,i} + d_{i,i}))^2 = 0. \quad (1)$$

So this term is always zero and to find  $\boldsymbol{\theta}$  only the remaining terms are used

$$\boldsymbol{\theta} = \arg \min_{\boldsymbol{\theta}} \sum_{\substack{1 \leq j \leq M \\ 1 \leq k \leq M \\ (j,k) \neq (i,i)}} (\hat{r}_{j,k} - (r(\boldsymbol{\theta})_{j,k} + d_{j,k}))^2, \quad (2)$$

Whether  $\hat{r}_{i,i}$  is available or not does not matter for the estimation of  $\boldsymbol{\theta}$  because it does not appear in (2). Now consider the case that  $\hat{r}_{i,i}$  is missing. The term in (1) can not be evaluated so it needs to be removed from the least squares problem. The remaining problem is (2). Whether  $\hat{r}_{i,i}$  is known or not does not matter for the estimation of  $\boldsymbol{\theta}$  because it does not appear in (2).

So in solving  $\boldsymbol{\theta}$  using the least squares method the cases that either  $d_{i,i}$  is unknown or  $\hat{r}_{i,i}$  is unavailable or both are all the same. This suggests that the theoretical best attainable performance is the same for all cases. But is this really true? To answer this question we will evaluate the CRB for four cases differing in whether we consider  $d_{i,i}$  as known or unknown and whether  $\hat{r}_{i,i}$  is included in the data or not.

## IV. PRELIMINARIES FOR THE DERIVATION

This section contains a summary of the background information and techniques necessary to solve the problem. References are given to more extensive treatments of the material presented here.

### A. Cramer Rao Bound

The Cramer Rao Bound is a lower bound on the variance of an unbiased estimator. Here we will only give its definition. A more extensive treatment can be found in [9, Ch. 3].

To find the CRB we need to know the probability density function (pdf) of the data  $\mathbf{x}$  as function of the unknown parameters  $\boldsymbol{\theta}$ . We denote the pdf by  $p(\mathbf{x}; \boldsymbol{\theta})$ . The pdf must satisfy the regularity condition

$$\mathbb{E} \left[ \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} \right] = 0 \quad \text{for all } \theta_i$$

The entries of the Fisher Information Matrix are given by

$$\begin{aligned} [\mathbf{F}(\boldsymbol{\theta})]_{i,j} &= -\mathbb{E} \left[ \frac{\partial^2 \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \\ &= -\mathbb{E} \left[ \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \ln p(\mathbf{x}; \boldsymbol{\theta})}{\partial \theta_j} \right]. \end{aligned}$$

Let  $\hat{\boldsymbol{\theta}}$  be an unbiased estimate of  $\boldsymbol{\theta}$  based on  $\mathbf{x}$ . The lower bound on the variance of  $\hat{\boldsymbol{\theta}}$  is given by the diagonal entries of  $\mathbf{F}$ .

### B. PDF and discarded data points

Consider the case that the pdf of the observations  $x_1 \dots x_n$  is given but the estimator uses only the observations  $x_1 \dots x_{n-1}$  (observation  $x_n$  is discarded/not available). To calculate the CRB the pdf of  $x_1 \dots x_{n-1}$  is needed. It can be found by integrating  $p(x_1, \dots, x_n)$  over  $x_n = -\infty \dots \infty$ , (see [10, Ch. 4, p. 202])

$$p(x_1, \dots, x_{n-1}) = \int_{-\infty}^{\infty} p(x_1, \dots, x_n) dx_n.$$

### C. Additional parameters

Assume that for a certain pdf  $p(\mathbf{x}; \boldsymbol{\theta})$  the Fisher Information Matrix is given by  $\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ , where

$$\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_m \end{bmatrix}.$$

The CRB is the inverse of  $\mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}}$ . Now the same data model is used but a parameter that first was considered known is considered unknown. The vector

of unknown parameters  $\boldsymbol{\theta}$  is extended by one entry  $\theta_{m+1}$  forming

$$\tilde{\boldsymbol{\theta}} = \begin{bmatrix} \boldsymbol{\theta} \\ \theta_{m+1} \end{bmatrix}.$$

The Fisher Information Matrix is extended by one extra row and column

$$\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}} = \begin{bmatrix} \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} & \mathbf{F}_{\boldsymbol{\theta}\theta_{m+1}} \\ \mathbf{F}_{\theta_{m+1}\boldsymbol{\theta}} & \mathbf{F}_{\theta_{m+1}\theta_{m+1}} \end{bmatrix}.$$

The new CRB is the inverse of  $\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}$  which can be partitioned as follows

$$\mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} = \begin{bmatrix} \left[ \mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} \right]_{\boldsymbol{\theta}\boldsymbol{\theta}} & \left[ \mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} \right]_{\boldsymbol{\theta}\theta_{m+1}} \\ \left[ \mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} \right]_{\theta_{m+1}\boldsymbol{\theta}} & \left[ \mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} \right]_{\theta_{m+1}\theta_{m+1}} \end{bmatrix}$$

Using the Schur complement [11, p. 264] the upper left block can be written as

$$\left[ \mathbf{F}_{\tilde{\boldsymbol{\theta}}\tilde{\boldsymbol{\theta}}}^{-1} \right]_{\boldsymbol{\theta}\boldsymbol{\theta}} = \left( \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{F}_{\boldsymbol{\theta}\theta_{m+1}} \mathbf{F}_{\theta_{m+1}\theta_{m+1}}^{-1} \mathbf{F}_{\theta_{m+1}\boldsymbol{\theta}} \right)^{-1}$$

Now we define a modified Fisher Information Matrix

$$\tilde{\mathbf{F}}_{\boldsymbol{\theta}\boldsymbol{\theta}} := \mathbf{F}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{F}_{\boldsymbol{\theta}\theta_{m+1}} \mathbf{F}_{\theta_{m+1}\theta_{m+1}}^{-1} \mathbf{F}_{\theta_{m+1}\boldsymbol{\theta}}.$$

The CRB for the original parameter set  $\boldsymbol{\theta}$  in the presence of an additional parameter  $\theta_{m+1}$  is given by the inverse of the modified Fisher Information Matrix  $\tilde{\mathbf{F}}$ .

### D. Wishart Distribution

If the  $M \times 1$  vectors  $\mathbf{x}_j$  are i.i.d. with a complex normal distribution with zero mean and variance  $\boldsymbol{\Sigma}$ , i.e.  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Sigma})$  then the matrix

$$\mathbf{W} = \sum_{j=1}^N \mathbf{x}_j \mathbf{x}_j^H$$

is said to have a complex Wishart distribution with  $N$  degrees of freedom, i.e.  $\mathbf{W} \sim \mathcal{CW}(\boldsymbol{\Sigma}, N)$ . The probability density function of the complex Wishart distribution is given by (see [12, Ch. 4.2, p. 90])

$$p(\mathbf{W}; \boldsymbol{\Sigma}) = \frac{1}{c_{M,N}} \frac{1}{|\boldsymbol{\Sigma}|^N} e^{-\text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{W})} |\mathbf{W}|^{N-M}$$

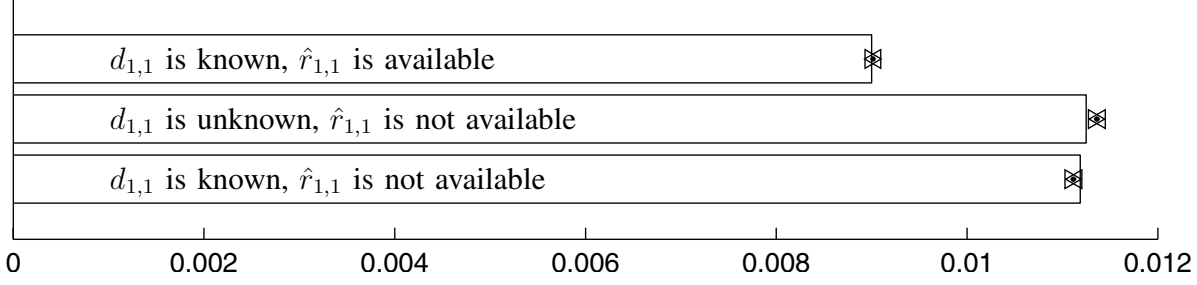


Fig. 1. Cramer-Rao bounds (bars) and Monte Carlo simulations of ML estimators including 99% confidence intervals (markers)

	$D_{i,i}$ known	$D_{i,i}$ unknown
$\hat{\mathbf{R}}_{i,i}$ included	$\mathbf{S} = \mathbf{Q}$	$\mathbf{S} = \mathbf{Q} - N\mathbf{P}$
$\hat{\mathbf{R}}_{i,i}$ excluded	$\mathbf{S} = \mathbf{Q} - (N - M)\mathbf{P}$	$\mathbf{S} = \mathbf{Q} - N\mathbf{P}$

TABLE I

THE FISHER INFORMATION MATRIX FOR THE FOUR DIFFERENT CASES IS GIVEN BY  $\mathbf{J}^H \mathbf{S} \mathbf{J}$ . THE MATRIX  $\mathbf{Q}$  IS DEFINED AS  $\mathbf{Q} = N (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1})$ .

This distribution has the following properties

$$\mathbb{E}[\mathbf{W}] = N\mathbf{\Sigma},$$

$$\begin{aligned} \text{cov}(\mathbf{W}) &= \mathbb{E}[\text{vec}(\mathbf{W})\text{vec}(\mathbf{W})^H] + \\ &\quad - \mathbb{E}[\text{vec}(\mathbf{W})]\mathbb{E}[\text{vec}(\mathbf{W})^H] \\ &= N\bar{\mathbf{\Sigma}} \otimes \mathbf{\Sigma}. \end{aligned}$$

## V. RESULTS

In this section the results summarized in table I will be derived.

From the Wishart distribution and the definition of the CRB one can derive the Fisher Information Matrix for Gaussian sources. The result is well known and given by Bang's formula

$$\mathbf{F} = \left( \frac{\partial \text{vec}(\mathbf{R})}{\partial \boldsymbol{\theta}^T} \right)^H N (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) \left( \frac{\partial \text{vec}(\mathbf{R})}{\partial \boldsymbol{\theta}^T} \right)$$

Now we consider the estimators that do not use diagonal element  $\hat{r}_{1,1}$ . The CRB for these estimators is based on the pdf of the remaining data. This pdf can be found by integrating the pdf of the Wishart distribution over  $\hat{r}_{1,1}$ . To present the result in compact matrix form we define  $\hat{\mathbf{R}}_0$  to be equal to  $\hat{\mathbf{R}}$  except for  $[\hat{\mathbf{R}}_0]_{1,1} = 0$ . We also define  $\hat{\mathbf{R}}_{1,1}$  as a submatrix of  $\hat{\mathbf{R}}$  obtained by omitting the first row

and the first column of  $\hat{\mathbf{R}}$ . We have found the pdf for  $\hat{\mathbf{R}}_0$  to be

$$\begin{aligned} p(\hat{\mathbf{R}}_0; \mathbf{R}) &= \frac{\Gamma(k+1)}{c_{m,n}} |\mathbf{R}|^{-N} |\hat{\mathbf{R}}_{1,1}|^{N-M} \times \\ &\quad ([\mathbf{R}^{-1}]_{1,1})^{N-M+1} e^{-\text{tr}(\mathbf{R}^{-1}\hat{\mathbf{R}}_0) + \frac{[\mathbf{R}^{-1}]_{1,1}}{[\hat{\mathbf{R}}_0^{-1}]_{1,1}}}. \end{aligned}$$

From this pdf the bounds can be derived. The derivation is somewhat tedious but the result can be expressed in a simple form. The Fisher Information Matrix is Bang's formula minus a 'penalty' for discarding data

$$\mathbf{F} = \mathbf{J}^H \left( N (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) - (N - M + 1)\mathbf{P} \right) \mathbf{J},$$

where  $\mathbf{J}$  is the Jacobian defined by

$$\mathbf{J} = \left( \frac{\partial \text{vec}(\mathbf{R})}{\partial \boldsymbol{\theta}^T} \right)$$

and  $\mathbf{P}$  is a 'penalty' matrix defined as

$$\mathbf{P} = \frac{\text{vec}(\mathbf{R}^{-1}\mathbf{E}_{1,1}\mathbf{R}^{-1})^H \text{vec}(\mathbf{R}^{-1}\mathbf{E}_{1,1}\mathbf{R}^{-1})}{[\mathbf{R}^{-1}]_{1,1}^2}.$$

If we extend the vector of unknowns to  $\tilde{\boldsymbol{\theta}}$ , it turns out that for both cases the Fisher Information Matrix is

$$\tilde{\mathbf{F}} = \mathbf{J}^H \left( N (\bar{\mathbf{R}}^{-1} \otimes \mathbf{R}^{-1}) - N\mathbf{P} \right) \mathbf{J}.$$

So, indeed if  $d_{1,1}$  is unknown the corresponding observation  $\hat{r}_{1,1}$  can be discarded without changing the bound. In other words, if  $d_{1,1}$  is unknown an algorithm that does not use  $\hat{r}_{1,1}$  can be efficient.

A more surprising result is that if  $\hat{r}_{1,1}$  has not been observed, it still matters whether  $d_{1,1}$  is known or not. To see whether it is possible to exploit the extra information we have implemented the Maximum Likelihood Estimators for three cases: 1)

$d_{1,1}$  is known,  $\hat{r}_{1,1}$  is available 2)  $d_{1,1}$  is unknown,  $\hat{r}_{1,1}$  is not available 3)  $d_{1,1}$  is known,  $\hat{r}_{1,1}$  is not available. We have run  $10^6$  Monte Carlo simulations for a scenario with the number of antennas  $M = 3$ , the number of samples  $N = 100$ . The parameters were the absolute gains and the phases. In figure 1 the CRB and the Mean Square Error of the absolute gain of the first element have been plotted. The figure shows that MLE performs better when  $d_{1,1}$  is known. The MSE lies even slightly below the bound. This is possible because the MLE for this problem is biased at low  $N$ . For high  $N$  the performance equals the bound, but also the differences between the two bounds goes to zero. In any practical scenario the number of samples  $N$  is far greater than the number of sensors  $M$ , and than the difference is negligible. Apart from this it is questionable there exist situations in practice where one knows  $d_{1,1}$ , but not  $\hat{r}_{1,1}$ .

## VI. CONCLUSIONS

Existing calibration algorithms for radio astronomical interferometers (or phased arrays) do not use the diagonal entries of the covariance estimate  $\hat{\mathbf{R}}$ . In case the noise powers are unknown these algorithms can still be efficient, since discarding the diagonal of  $\hat{\mathbf{R}}$  has no influence on the Cramer Rao Bound.

## 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 Widefield 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] A. J. Boonstra and A. J. van der Veen, "Gain Calibration Methods for Radio Telescope Arrays," *IEEE Trans. Signal Processing*, vol. 51, no. 1, pp. 25–38, Jan. 2003.
- [5] 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*, vol. 17, no. 1, pp. 35–42, 2004.
- [6] 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)*, Mar. 2005.
- [7] B. D. Jeffs, S. van der Tol and A. J. van der Veen, "Direction Independent Self Calibration of Large Distributed Sensor Arrays," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2006, accepted.
- [8] S. J. Wijnholds and A. J. Boonstra, "A multisource calibration method for phased array radio telescopes," in *Fourth IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, 2006, to be submitted.
- [9] S. Kay, *Fundamentals of statistical signal processing: Estimation theory*. PTR, Prentice Hall, 1993.
- [10] A. Leon-Garcia, *Probability and Random Processes for Electrical Engineering*, 2nd ed. Reading, MA: Addison-Wesley, 1994.
- [11] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Prentice Hall, 2000.
- [12] D. R. Brillinger, *Time Series: Data Analysis and Theory*. Holt, Rinehart and Winston, inc., 1975.