Autonomous Online LOFAR Station Calibration

Stefan J. Wijnholds[1]

[1]ASTRON, R&D department, Oude Hoogeveensedijk 4, P.O. Box 2, NL-7990 AA Dwingeloo, The Netherlands
Email: wijnholds@astron.nl

Abstract

The spatial filtering performance of the LOFAR station beam former strongly depends on the quality of the station calibration. The station calibration runs online to track slow variations in receiver system response over time. Since the LOFAR array will ultimately consist of well over 50 stations, the algorithm should run autonomously. Finally, the all-sky calibration problem at station level should be solved using the computational resources that can be provided by one or at most a few CPUs. This paper provides an overview of the station calibration data reduction pipeline and presents results from a demonstration on actual data.

1. Introduction

The LOFAR stations currently being rolled out in the Netherlands will consist of 48 low band antennas (LBAs) operating in the 10-90 MHz frequency range and 48 high band antenna (HBA) tiles operating between 120 and 240 MHz. The antennas as well as the on-site electronics are subject to gradual changes in the environmental conditions, such as humidity and temperature, that may introduce variations in the response of the analog system (antennas, cables and receivers). The station calibration should characterize these variations and apply appropriate modifications to the beam former coefficients to ensure optimal performance of the station beam former. This will ensure optimal spatial suppression of RFI and limit pointing errors, both reducing the complexity of the data reduction of the interstation visibilities.

This context poses a number of interesting challenges. Firstly, with over 50 stations, the station calibration should be an autonomous process that can run without continuous monitoring by a human operator. Secondly, since the antennas are sensitive to every direction on the sky, the station calibration is effectively all-sky calibration, i.e. in this context, wide field means horizon to horizon. Thirdly, the station calibration should thus be robust to this harsh and possibly changing RFI environment. Finally, the station calibration process should not require more processing power than can be provided by one or at most a few CPUs due to physical and financial constraints.

This paper describes the station calibration pipeline. The next section gives an overview of this pipeline and describes the data model used in the rest of this paper. Sections 3 and 4 discuss the RFI detection and calibration algorithms respectively. The resulting calibration corrections should be checked carefully before applying them to the station beam former, because erroneous results will cause the beam former to utterly fail. This important step is treated in section 5. Section 6 presents an end-to-end demonstration of the complete pipeline just before this paper is concluded in section 7.

Notation: The transpose operator is denoted by ^{T}, the complex conjugate (Hermitian) transpose by ^{H} and the pseudo-inverse by †. An estimated value is denoted by $\mathcal{E}\{\cdot\}$, $\mathrm{diag}(\cdot)$ converts a vector to a diagonal matrix with the vector placed on the main diagonal and $\mathrm{vec}(\cdot)$ converts a matrix to a vector by stacking the columns of the matrix.

2. Overview of the station calibration pipeline and data model

Each LBA or HBA tile in a LOFAR station is connected to two receiver units, one for each polarization. The receiver unit is equipped with appropriate analog filters and an analog-to-digital converter which samples the signal at a rate

of either 160 or 200 MHz. The signals from the receiver units are fed into a polyphase filter bank providing 512 subbands, which are either 156 or 195 kHz wide. A selection of eight 4 MHz blocks of subbands is then fed into the beamformer which applies appropriate phase and amplitude corrections to the subband data to provide up to eight beams with a total bandwidth of 32 MHz. The FX-type station correlator running parallel to the beam former can correlate all incoming signals for a single subband in real time thus providing an array covariance matrix (matrix of visibilities) for that subband.

The station calibration uses the output from the station correlator as input and assumes a sky model based on celestial sources. The array covariance matrices should thus be RFI free to produce correct results. Although the calibration results are checked before they are used in the station beam former, RFI detection is computationally cheaper than calibration (order p^2 versus p^3 where p is the number of input signals which can be order 100 for a LOFAR station) and thus save computational resources.

The RFI detection and calibration algorithms are described hereafter using the following data model. Let the output signal of the i^{th} element be denoted by $x_i(t)$ and define the array signal vector $\mathbf{x}(t) = [x_1(t), x_2(t), \cdots, x_p(t)]^T$. We assume the presence of q mutually independent i.i.d. (independent (in time) identically distributed) Gaussian signals $s_k(t)$ impinging on the array, which are stacked in a $q \times 1$ vector $\mathbf{s}(t)$. Likewise, the sensor noise signals $n_i(t)$ are assumed to be mutually independent i.i.d. Gaussian signals and are stacked in a $p \times 1$ vector $\mathbf{n}(t)$. If the narrow band condition holds, we can define the q spatial signature vectors $\mathbf{a}_k$, which include the phase delays due to the geometry and the directional response of the antennas. They are assumed to be known. The direction-independent complex element gains can be described as $\mathbf{g} = [g_1, g_2, \cdots, g_p]^T$ with corresponding diagonal matrix form $\mathbf{G} = \mathrm{diag}(\mathbf{g})$. With these definitions, the array signal vector can be described as

$$\mathbf{x}(t) = \mathbf{G}\left(\sum_{k=1}^{q} \mathbf{a}_k s_k(t)\right) + \mathbf{n}(t) = \mathbf{G}\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t), \tag{1}$$

where $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_q]$ (size $p \times q$). The signal is sampled with period T and N sample vectors are stacked in a data matrix $\mathbf{X} = [\mathbf{x}(T), \cdots, \mathbf{x}(NT)]$. The covariance matrix of $\mathbf{x}(t)$ is $\mathbf{R} = \mathcal{E}\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\}$ and is estimated by $\widehat{\mathbf{R}} = N^{-1}\mathbf{X}\mathbf{X}^H$. Likewise, the source signal covariance $\mathbf{\Sigma}_s = \mathrm{diag}(\boldsymbol{\sigma}_s)$ where $\boldsymbol{\sigma}_s = [\sigma^2_{s_1}, \sigma^2_{s_2}, \cdots \sigma^2_{s_q}]^T$ is a vector of source powers, and the noise covariance matrix is $\mathbf{\Sigma}_n = \mathrm{diag}(\boldsymbol{\sigma}_n)$ where $\boldsymbol{\sigma}_n = [\sigma^2_{n_1}, \sigma^2_{n_2}, \cdots \sigma^2_{n_p}]^T$ is a vector of system noise powers. Then the model for $\mathbf{R}$ based on (1) is

$$\mathbf{R} = \mathbf{G}\mathbf{A}\mathbf{\Sigma}_s\mathbf{A}^H\mathbf{G}^H + \mathbf{\Sigma}_n\,. \tag{2}$$

3. RFI detection

Leshem and Van der Veen [1] have derived a generalized likelihood ratio test (GLRT) for the detection of RFI under the assumption that the power of the astronomical signals is negligible compared to the system noise and the RFI power. They also show that a simple ad hoc test based on the Frobenius norm of the whitened array covariance matrix has the same asymptotic performance as the GLRT. In sky noise dominated systems like LOFAR [2], the power of astronomical sources like the galactic plane and the north polar spur is not negligible compared to the sky noise. Furthermore, these structures become more and more resolved towards higher frequencies in the band. This aspect is handled by comparing the whitened array covariance matrices of consecutive subbands, i.e. by using the test

$$t \equiv \max\left(\| \widehat{\mathbf{R}}_w^n \|_F - \| \widehat{\mathbf{R}}_w^{n-1} \|_F, \| \widehat{\mathbf{R}}_w^n \|_F - \| \widehat{\mathbf{R}}_w^{n+1} \|_F\right) \gtrless \gamma_{RFI}, \tag{3}$$

where the superscripts denote the subband indices. The RFI detection threshold γ_{RFI} was set quite loosely on 5σ above the expected noise on $\| \widehat{\mathbf{R}} \|_F$. This ensures a low false rejection rate to ensure maximal use of the available data in the rest of the pipeline.

4. Calibration

The goal of the station calibration is to determine the direction independent complex gains of the receiver paths. A processing cost effective and asymptotically statistically efficient solution to this problem is presented and discussed

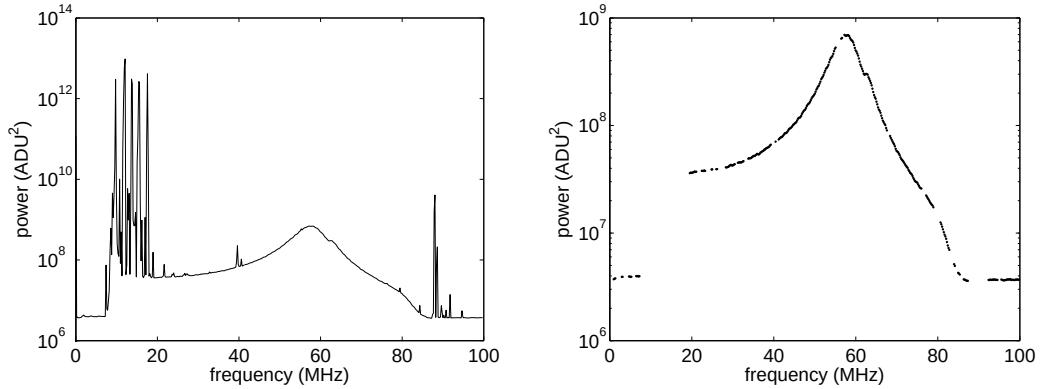

Fig. 1. A typical LBA spectrum before (left) and after (right) blanking of RFI occupied subbands.

in [3]. In this paper I only state the result, which is based on the observation that $g_i\overline{g}_k = \widehat{R}_{ik}/R_{0,ik}$ holds for all off-diagonal elements of $\widehat{\mathbf{R}}$ and the model visibilities $\mathbf{R}_0 = \mathbf{A}\mathbf{\Sigma}_s\mathbf{A}^H$, i.e. for $i \neq k$. Since the index k can be chosen freely as long as $k \neq i, j$, we can introduce $\mathbf{c}_{1,ij}$ being the column vector containing the values $\widehat{R}_{ik}R_{0,jk}$ and $\mathbf{c}_{2,ij}$ being the column vector containing the values $R_{0,ik}\widehat{R}_{jk}$ for all possible values of k. The ratio between complex gains g_i and g_j can now be computed by

$$\frac{g_i}{g_j} = \mathbf{c}_{2,ij}^{\dagger}\mathbf{c}_{1,ij}. \tag{4}$$

These gain ratios can be collected in a matrix $\mathbf{M}$ with entries $M_{ij} = g_i/g_j$. $\mathbf{g}$ can be extracted from this matrix using an EVD. This approach allows us to neglect the unknown receiver noise powers and apply baseline restrictions by putting appropriate constraints on the values of k.

The quotient g_i/g_j is insensitive to modification by a constant scaling factor α applied to all gains. This factor α can be determined by introducing the vectors $\mathbf{r}_0 = \mathrm{vec}_{-}(\mathbf{G}\mathbf{R}_0\mathbf{G}^H)$ and $\mathbf{r} = \mathrm{vec}_{-}(\widehat{\mathbf{R}})$, where $\mathrm{vec}_{-}(\cdot)$ operates like the $\mathrm{vec}(\cdot)$ operator but leaves out the elements on the main diagonal of its argument and computing $\widehat{\alpha} = \sqrt{\mathbf{r}_0^{\dagger}\mathbf{r}}$.

5. Handling erroneous calibration results

Since the data captured by the correlator from consecutive subbands is acquired only 1 s apart, the state of the instrument and the sky has normally hardly changed. The response of the receiving system is also expected to change only gradually with frequency. This implies that the gain solutions obtained from the calibration algorithm are expected to behave smoothly over frequency. We can therefore define the following test to assess the calibration result:

$$\mathbf{t} \equiv \max\left(\left|\mathbf{g}^n - \mathbf{g}^{n-1}\right|, \left|\mathbf{g}^n - \mathbf{g}^{n+1}\right|\right) \gtrless \gamma\mathbf{1}, \tag{5}$$

where $\mathbf{1}$ denotes a vector of appropriate size filled with ones. Note that this test is applied to all calibrated signal paths individually. This allows, in principle, to flag the gain solutions per subband and per receiving element.

At the end of the pipeline, we only have valid calibration results for a limited number of subbands. The other subbands either contained RFI or did not produce correct results for some reason, for example due to low power RFI that escaped detection. The correction factors for those subbands can be estimated by fitting a band pass model to the available results to inter- and extrapolate the results to all subbands.

6. Demonstration on actual LOFAR data

In this section, the data reduction pipeline described in the previous sections is applied to a data set that was taken with a 48 element LBA station in Exloo (The Netherlands). The measurement was started on 18 January 2008 at 8:23:58

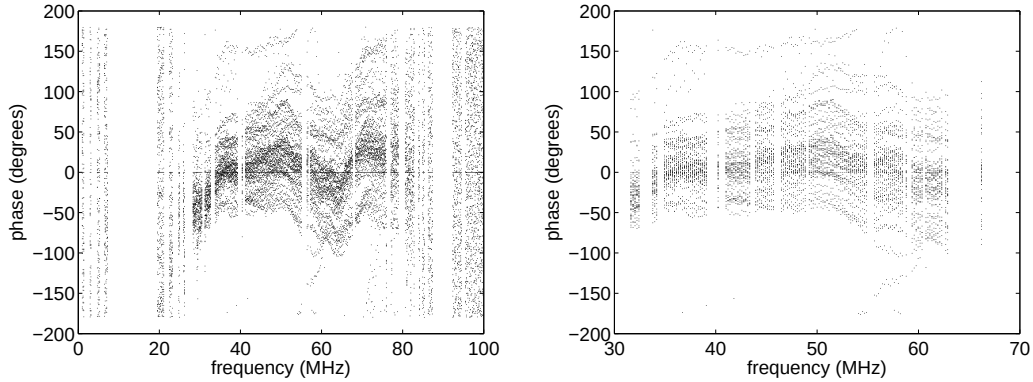

Fig. 2. The phase of the gain solutions before (left) and after (right) blanking of erroneous results.

UTC and lasted for 512 s scanning all 512 195 kHz subbands integrating 1 s on each subband. The left panel of Fig. 1 shows a typical LBA autocorrelation spectrum showing the 10-90 MHz band pass filter and RFI in the low ($<$25 MHz) and high ($>$88 MHz) frequency ends of the observing band. The autocorrelation spectrum after blanking is shown in the right panel of Fig. 1.

A tolerant setting of the RFI detection threshold increases the probability that low power RFI escapes detection. The consequences are demonstrated clearly in the left panel of Fig. 2 showing the phase of the gain solutions. Note that one element has been used as phase reference. Therefore the phase behavior shown in this plot is the difference in phase behavior between two receiving paths showing, amongst others, the effect of the antenna resonance. The detection threshold per element was set to a relative error of 20%. If at least 40 out of 48 antennas satisfied this, the data was regarded as good. This data is shown in the right panel of Fig. 2. This threshold is an experimental tradeoff for this particular measurement. It is therefore set as loose as possible to allow as much reliably looking data as possible to get through. In an autonomous system, the threshold will be set by a combination of experience based on monitoring data and statistical analysis and will be more stringent. As a result, the outermost frequency points will no longer be available anymore showing the importance of a good band pass description to ensure correct extrapolation to the lower and upper frequencies in the observing band.

7. Conclusions

The autonomous calibration works on actual data given the available sensitivity and appropriate thresholds for detection of RFI and erroneous results. The ability to run autonomously implies that the test statistic used to check the validity of the calibration result will be subjected to a very stringent threshold to ensure a very low false acceptance rate. If this step works well in practice, we can impose a very loose threshold in the RFI detection, to get a low false rejection rate ensuring maximal use of the available data.

8. References

[1] Amir Leshem and Alle-Jan Van der Veen, "Multichannel Detection of Gaussian Signals with Uncalibrated Receivers," *IEEE Signal Processing Letters*, vol. 8, no. 4, pp. 120–122, Apr. 2001.

[2] Stefan J. Wijnholds, Jaap D. Bregman, and Albert-Jan Boonstra, "Sky Noise Limited Snapshot Imaging in the Presence of RFI with LOFAR's Initial Test Station," *Experimental Astronomy*, vol. 17, no. 1-3, pp. 35–42, 2004.

[3] Stefan J. Wijnholds and Albert-Jan Boonstra, "A Multisource Calibration Method for Phased Array Radio Telescopes," in *Fourth IEEE Workshop on Sensor Array and Multi-channel Processing (SAM)*, Waltham (MA), 12-14 July 2006.