Statistical analysis of multi-source Calibration

S. Kazemi

Kapteyn Astronomical Institute
kazemi@astro.rug.nl

Presentation in CALIM, August 2010
(Under supervision of S. Zaroubi, and S.Yatawatta)

Introduction

The scientific goals of new radio arrays (e.g. LOFAR & SKA) require extreme sensitivity and dynamic range (e.g. LOFAR EOR)

Calibration

Estimating unknown instrument and the sky parameters and correcting them before imaging.

Calibration challenges:

- Accuracy of calibration algorithms becomes crucial for achievement of scientific goals
- The large number of data requires fast algorithms

Input noise
"Instrumental noise"
"Sky noise"
"RFI"

→ **Calibration process** →

output noise
"Input noise"
+
Noise of the calibration algorithm
"solver noise"

The aim is to devise new calibration methods which minimize the solver noise

Outline

www.lofar.org/

2 Measurement equation

3 The LS, EM, and SAGE algorithms

4 Solver noise

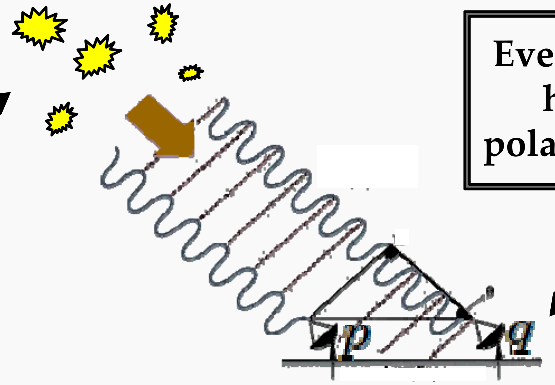
5 Illustrative example

6 Conclusions and future work

Measurement equation

Initial assumptions

Radio frequency sky is decomposed to K separated sources far away from the array



Every antenna has dual polarized feeds X Y

Baseline pq

Calibration problem modeling

Hamaker-Bregman-Sault, 1996.

Jones matrix: describes amplifier gains, beam shapes, ionospheric effects, and etc

Coherency: describes polarization state of the source i

$$\begin{bmatrix} XX & XY \\ YX & YY \end{bmatrix}_{pq} = \mathbf{V}_{pq} = \sum_{i=1}^{K} \mathbf{J}_{pi}(\boldsymbol{\theta})\mathbf{C}_i\mathbf{J}_{qi}^H(\boldsymbol{\theta}) + \mathbf{N}_{pq}$$

Measured visibility

Noise matrix of the baseline pq

Vectorized form

$$\mathbf{y} = \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta}) + \mathbf{n}, \quad \mathbf{n} \sim \mathcal{N}(0, \boldsymbol{\Pi})$$

The LS, EM, and SAGE algorithms

The Least Squares (LS, Normal) calibration method (AIPS, AIPS++, CASA)

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} ||\mathbf{y} - \sum_{i=1}^{K} \mathbf{s}_i(\boldsymbol{\theta})||^2 \xrightarrow[\text{technique}]{\text{Levenberg Marquardt}} \boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - (\nabla_{\boldsymbol{\theta}}\nabla_{\boldsymbol{\theta}}^T \phi(\boldsymbol{\theta}) + \lambda \mathbf{H})^{-1} \nabla_{\boldsymbol{\theta}} \phi(\boldsymbol{\theta})|_{\boldsymbol{\theta}^k}$$

•Easy to program

But,

•Heavy computational cost of $O((KN)^2)$

•Slow rate of convergence

Feder and Weinstein, 1988.

Expectation Maximization (EM) algorithm

Y : Observed data

X : Complete data

$\boldsymbol{\theta}^*$: Parameter value

$$\log f_{\mathbf{Y}}(\mathbf{y};\boldsymbol{\theta}) = E\{\log f_{\mathbf{X}}(\mathbf{x};\boldsymbol{\theta})|\mathbf{Y}=\mathbf{y};\boldsymbol{\theta}^*\} - E\{\log f_{\mathbf{X}|\mathbf{Y}=\mathbf{y}}(\mathbf{x};\boldsymbol{\theta})|\mathbf{Y}=\mathbf{y};\boldsymbol{\theta}^*\}$$

$\boxed{E_{\mathbf{X}|\mathbf{Y}=\mathbf{y},\boldsymbol{\theta}^*}\{\log L(\boldsymbol{\theta})\}}$

Jensen's inequality: this expected value will be decreased for all $\boldsymbol{\theta} \neq \boldsymbol{\theta}^*$

Expectation-Step: **Compute the expected value of the log-likelihood function, with respect to conditional distribution of the complete data X given observed data y under the current parameter estimate θ^k**

Maximization-Step: **Maximize the log-likelihood with respect to θ**

Yatawatta et al. 2009, Kazemi et al. in prep.

Assuming that the contribution of source *i* depends only on a subset of parameters θ_i and partitioning the parameter vector as $\theta = [\theta_1^T \ \theta_2^T \ldots \theta_K^T]^T$,

$$\mathbf{x}_i = \mathbf{s}_i(\theta_i) + \mathbf{n}_i, \quad for\ i \in \{1, 2, \ldots K\}$$

$$\mathbf{n}_i \sim \mathcal{N}(0, \beta_i \Pi)$$

$$\longrightarrow \quad \boxed{\mathbf{y} = \sum_{i=1}^{K} \mathbf{x}_i} \quad \longrightarrow$$

E- Step: $\widehat{\mathbf{x}}_i^k = E\{\mathbf{x}_i | \mathbf{y}, \theta^k\}$

M- Step: $\min_{\theta_i} \ \phi_i(\theta_i) = ||[\widehat{\mathbf{x}}_i^k - \mathbf{s}_i(\theta_i)](\beta_i \Pi)^{-\frac{1}{2}}||^2$

Fessler and Herro, 1994.

Space Alternating Generalized Expectation Maximization (SAGE) algorithm

- **Dedicating whole the noise to only one source**
- **Utilizing the EM algorithm**

Yatawatta et al. 2009, Kazemi et al. in prep.

$$\mathbf{x}^i = \mathbf{s}_i(\boldsymbol{\theta}_i) + \mathbf{n} \;\Longrightarrow\; \boxed{\mathbf{y} = \mathbf{x}^i + \sum_{\substack{l=1\\ l\neq i}}^{K} \mathbf{s}_l(\boldsymbol{\theta}_l)} \;\Longrightarrow$$

E- Step: $\widehat{\mathbf{x}}_i^k = E\{\mathbf{x}_i|\mathbf{y},\boldsymbol{\theta}^k\}$

M- Step: $\min_{\boldsymbol{\theta}_i} \phi_i(\boldsymbol{\theta}_i) = ||[\widehat{\mathbf{x}^i}^k - \mathbf{s}_i(\boldsymbol{\theta}_i)](\boldsymbol{\Pi})^{-\frac{1}{2}}||^2$

Advantages of the EM algorithm over the Normal algorithm:

- Breaking the likelihood maximization into smaller computational steps
- Computational cost equal to $KO(N^2)$

$$\boldsymbol{\theta}_i^{k+1} = \boldsymbol{\theta}_i^k - (\nabla_{\boldsymbol{\theta}_i}\nabla_{\boldsymbol{\theta}_i}^T \phi_i(\boldsymbol{\theta}_i) + \lambda \mathbf{H}_i)^{-1} \nabla_{\boldsymbol{\theta}_i}\phi_i(\boldsymbol{\theta}_i)|_{\boldsymbol{\theta}_i^k} \quad \text{for } i \in \{1,2,\ldots,K\}$$

- Increasing the likelihood at each iteration step
- Improving the speed of convergence compared with LS calibration

But still,

- Possibility of converging to a local optimum
- Implementation of the algorithm is complicated compared with LS method

Additional advantages of using the SAGE algorithm:

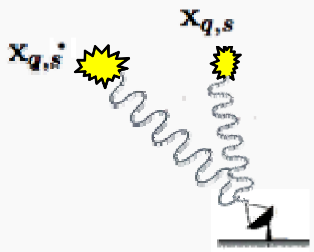
- Improving the speed of convergence
- Improving the accuracy of calibration results

Solver noise

Initial assumptions

The goal is to minimize the "distance" between the real gains and calculated solutions = Minimizing the solver noise

$x_{q,s}$:= gain solutions for the source *s* at the antenna *q*



- Sky noise for two different directions is uncorrelated
- Thermal noise for any number of directions from the same station is the same
- Solver noise is the same for all stations and directions

Kullback-Leibler Divergence (KLD)

$$KLD(f,g) = \sum_{\mathbf{x}} f(\mathbf{x}) log \frac{f(\mathbf{x})}{g(\mathbf{x})}$$

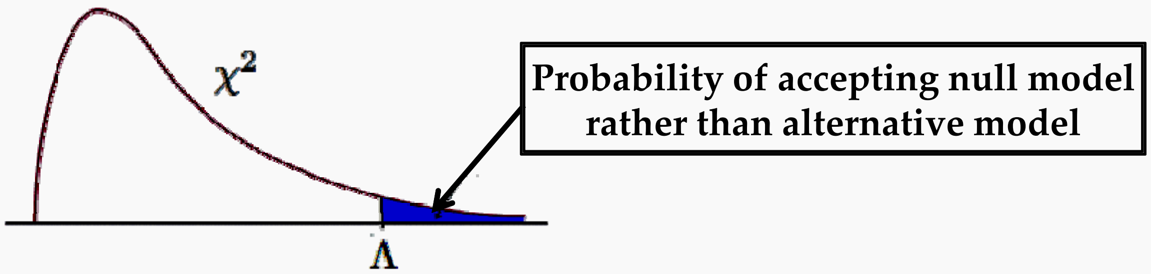
The higher KLD between two different sources solutions at the same antenna, the solver noise is less

- We fit a Gaussian mixture model to the solutions by EM algorithm
- KLD is calculated via Monte-Carlo algorithm

Likelihood Ratio Test (LRT)

Null model: Independent solutions for different sources for the same antenna
Alternative model: Dependent solutions for different sources for the same antenna

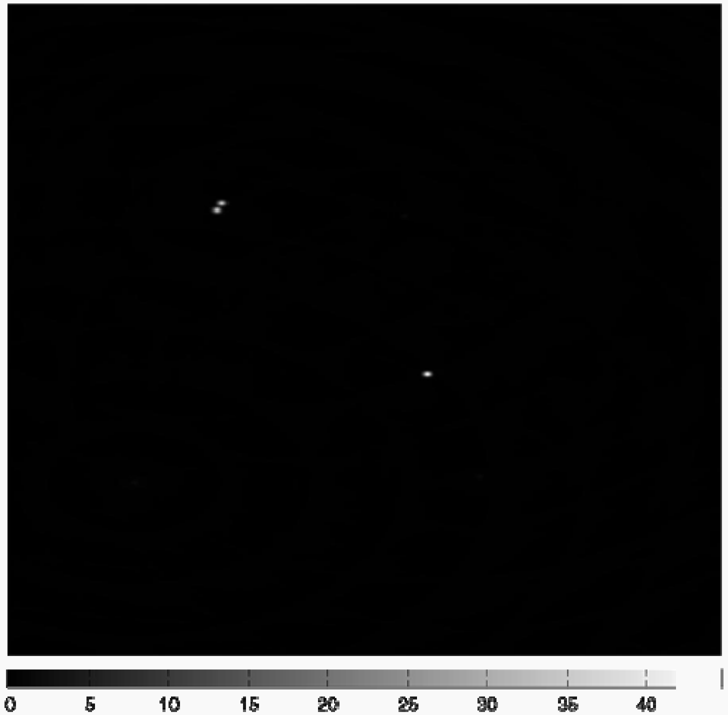
$$\Lambda = -2\ln\left(\frac{\text{Likelihood for null model}}{\text{Likelihood for alternative model}}\right)$$



The less LRT between two different sources solutions at the same antenna, the solver noise is less

Illustrative example

One channel simulated observation of three sources by fourteen antennas



Clean image

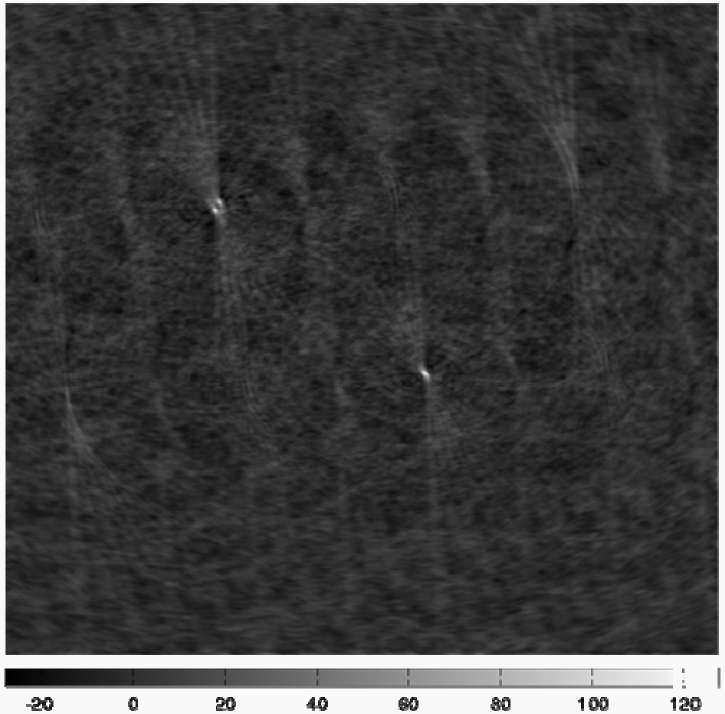
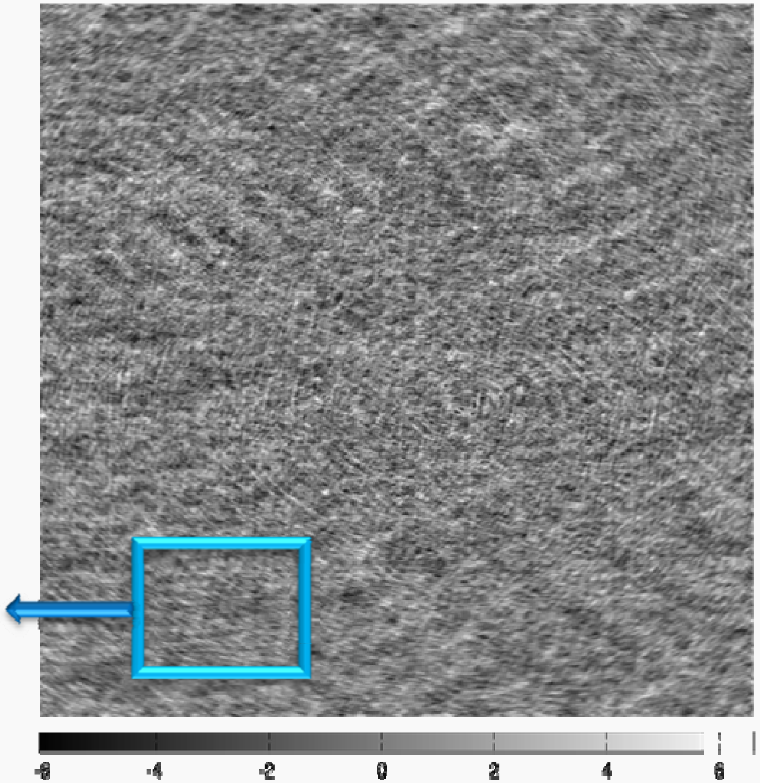


Image having gain errors and white Gaussian noise

Residual Errors

SAGE, 9 iterations

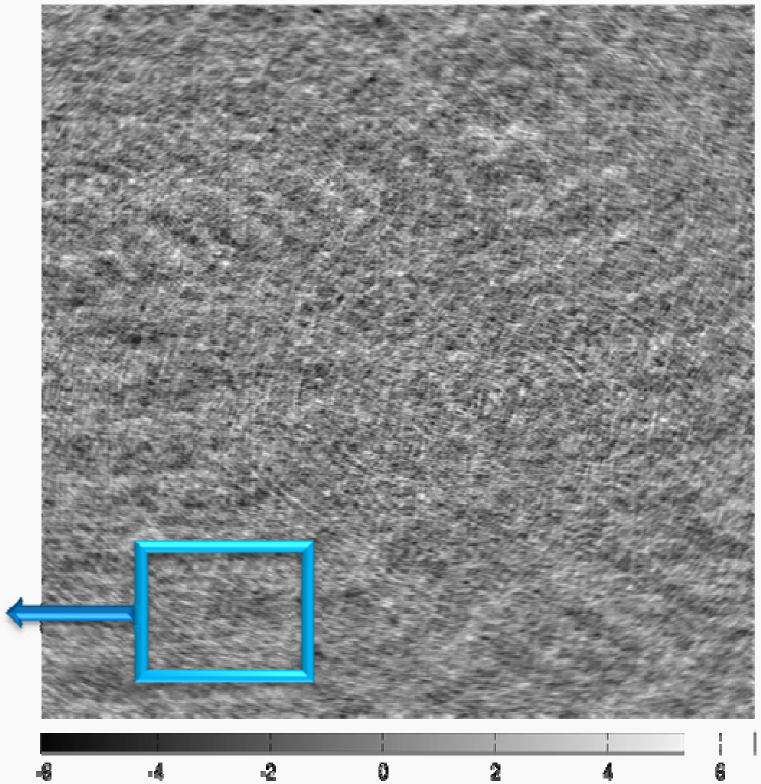
STD=1.2179



RMSE= 3.9259

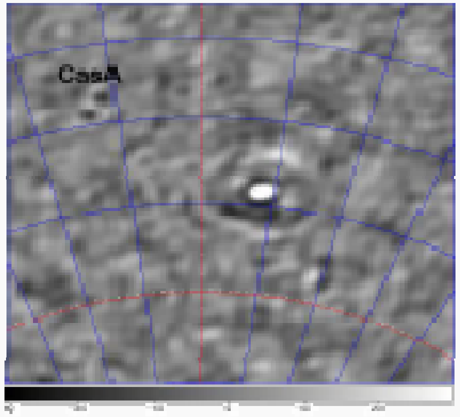
NORMAL, 9 iterations

STD=1.2186

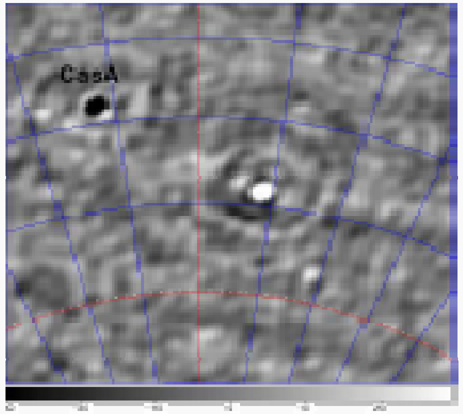


RMSE= 3.9505

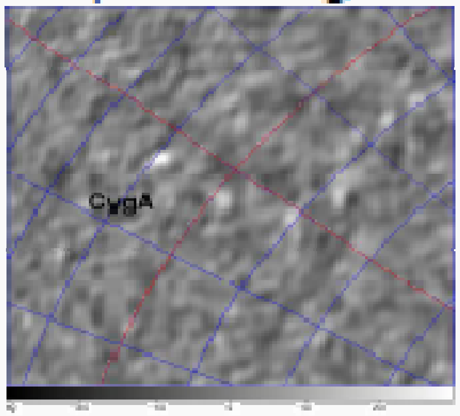
LOFAR CS1 images from CasA and CygA



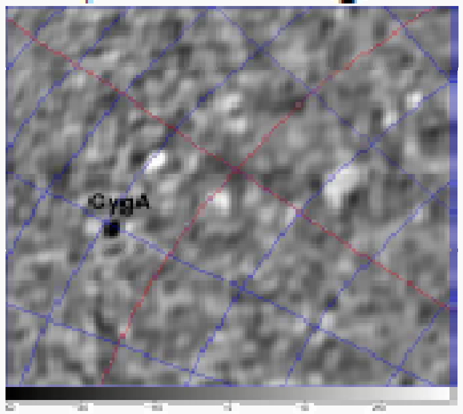
CasA, SAGE Algorithm



CasA, Normal Algorithm



CygA, SAGE Algorithm



CygA, Normal Algorithm

Yatawatta et al. 2009.

Conclusions and future work

Conclusions:

- The SAGE algorithm is superior in terms of accuracy and speed of convergence
- The non-linear problem requires suitable regularization
- The KLD and the LRT could be used to study the influence of solver noise

Future work:

- Apply the algorithms to real data of LOFAR
- Study different regularization methods
- Implement different noise models


Thank you

Illustrative example

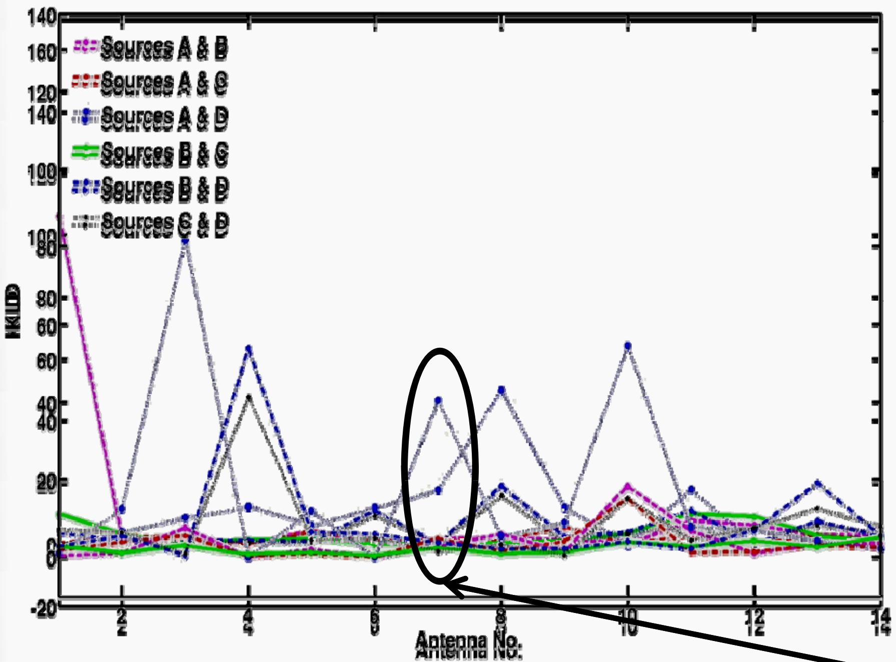
rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Detection solver noise via KLD and LRT

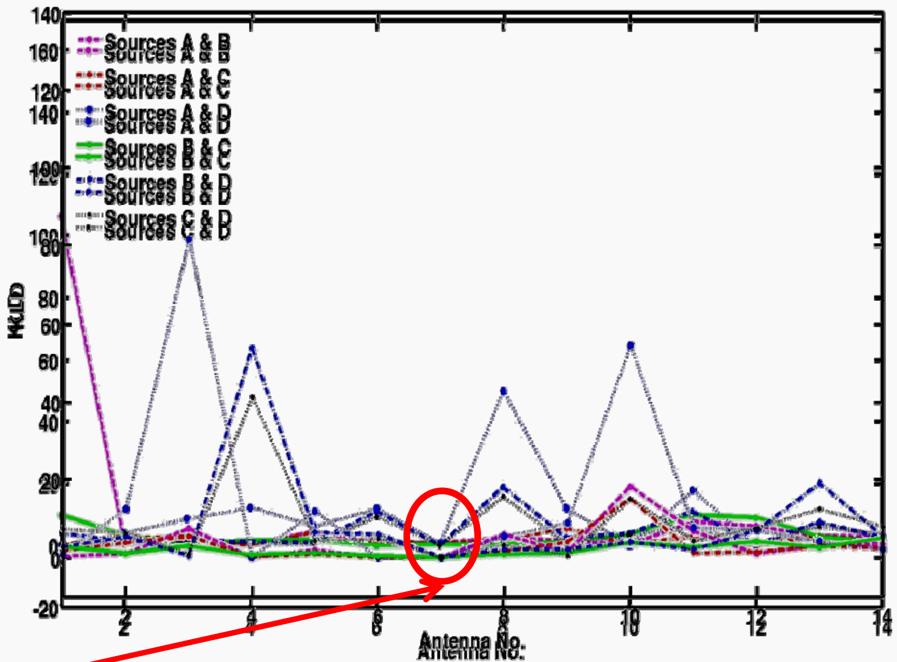
The SAGE calibration
No calibration

Solutions of antenna seven are given *sin(t)* noise where *t* is integration time

Real Solutions



Solution added *sin(t)* noise



The KLD of the noisy antenna becomes zero

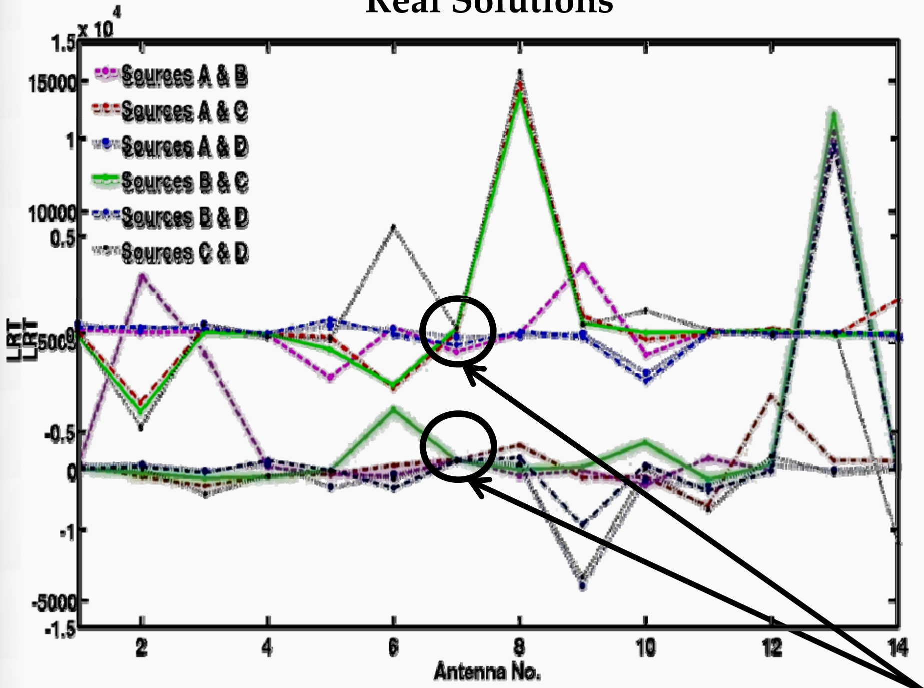
Illustrative example

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

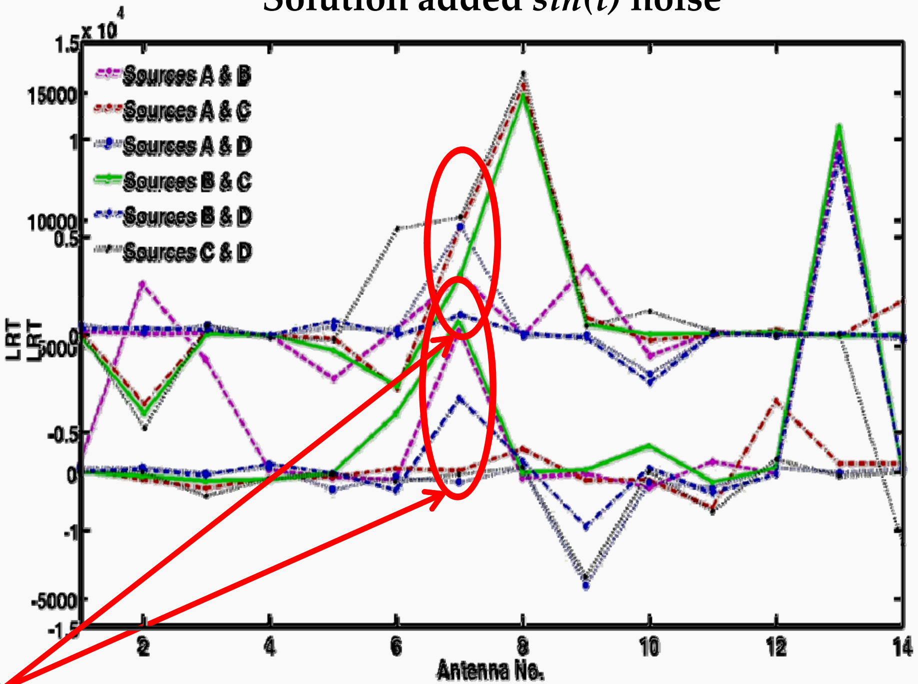
Detection solver noise via KLD and LRT

The SAGE calibration
No calibration

Real Solutions



Solution added *sin(t)* noise



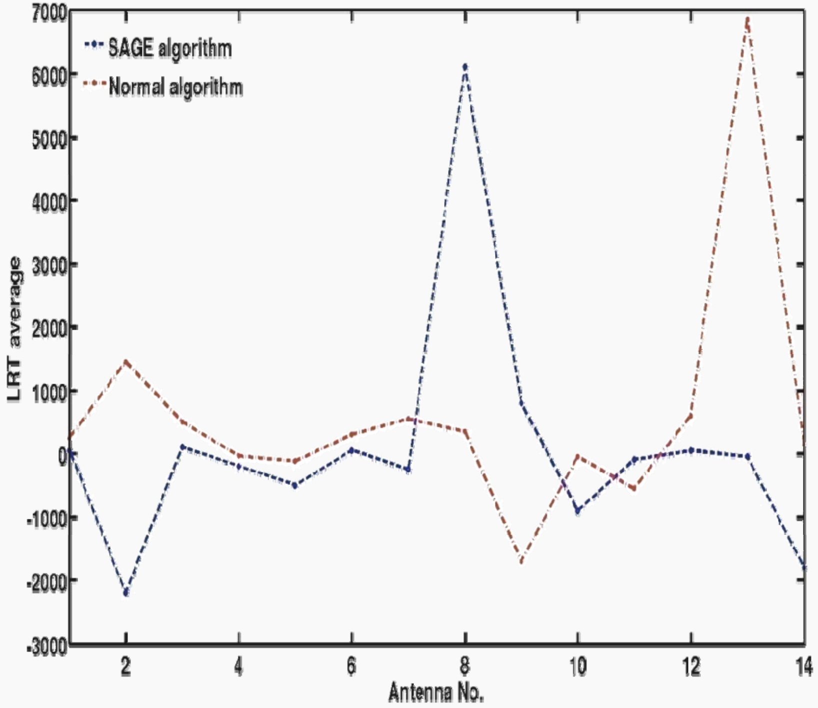
The LRT of the noisy antenna becomes higher

Illustrative example

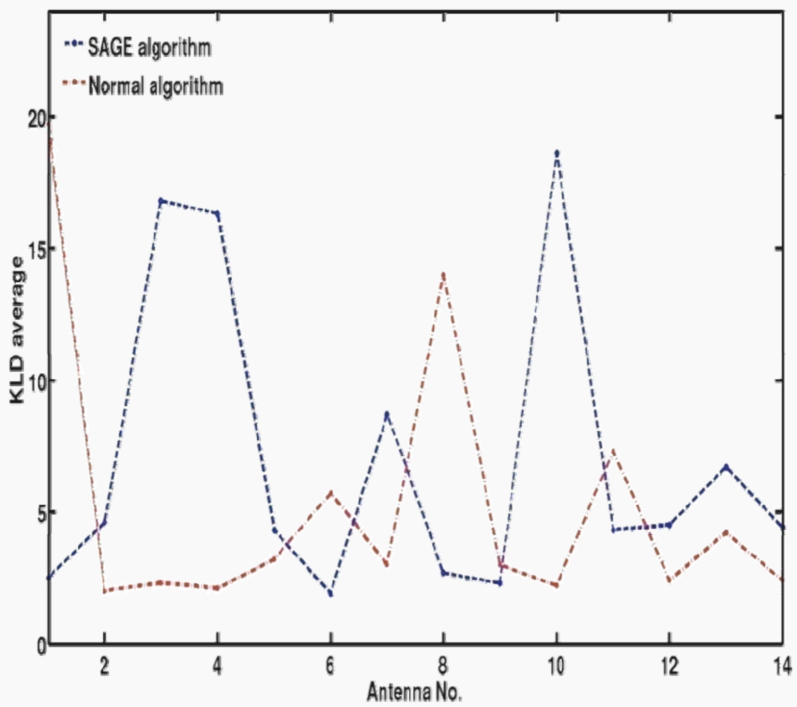
rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

The Normal and the SAGE algorithms with 16 number of iterations

KLD comparison



LRT comparison

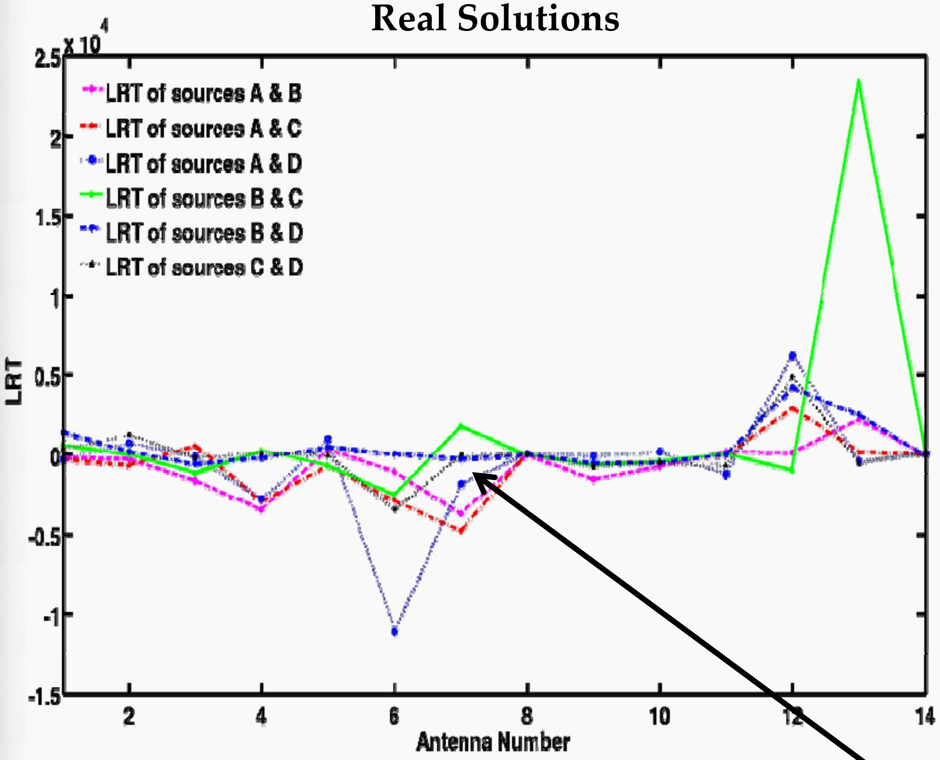


The SAGE algorithm solutions have superior accuracy

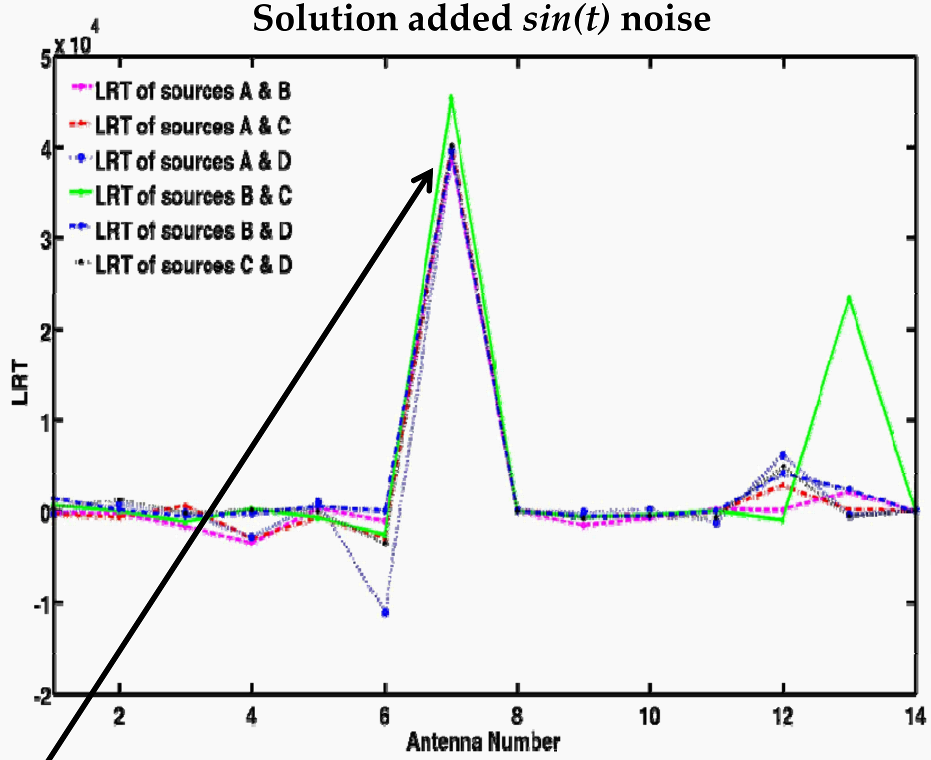
Illustrative example

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Real Solutions



Solution added *sin(t)* noise

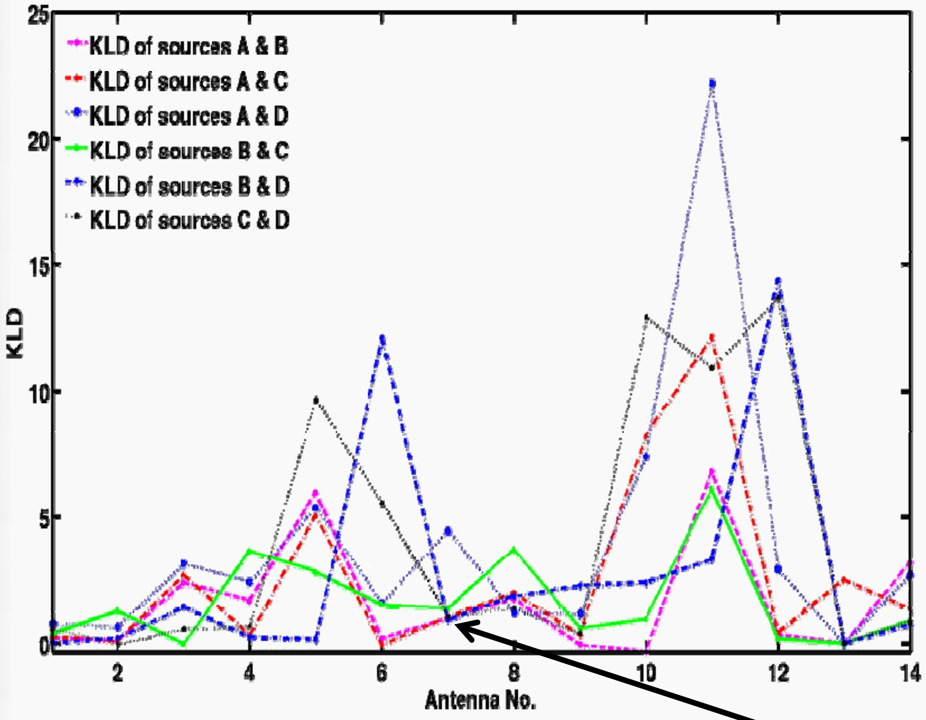


The LRT of noisy antenna is highly increased

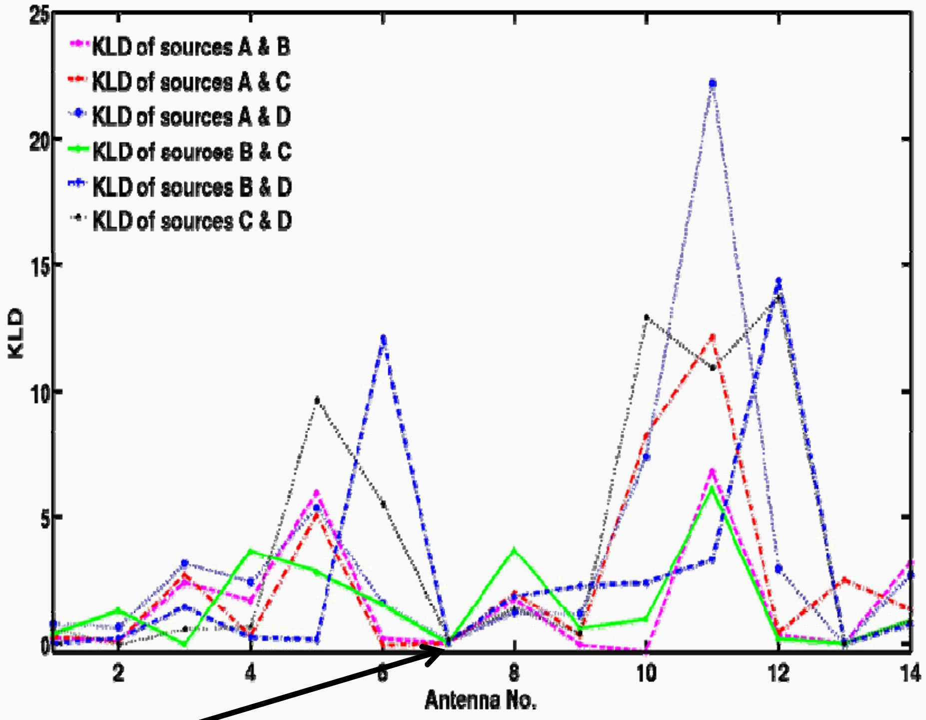
Illustrative example

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Real Solutions



Solution added *sin(t)* noise



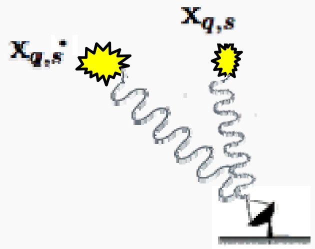
The KLD of noisy antenna becomes almost zero

Noise in solutions

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Initial assumptions

$$\mathbf{J}_{q,s} = \begin{bmatrix} J_{11,q} & 0 \\ 0 & J_{22,q} \end{bmatrix}_s \qquad \mathbf{x}_{q,s} = [\Re e(J_{11,q})\ \Im m(J_{11,q})\ \Re e(J_{22,q})\ \Im m(J_{22,q})]_s^T$$



- Sky noise for two different directions is uncorrelated
- Thermal noise for any number of directions from the same station is the same
- Solver noise is the same for all stations and directions

The goal is to minimize the "distance" between the real gains and calculated solutions = Minimizing the solver noise

Kullback-Leibler Divergence (KLD)

$$KLD(f,g) = \sum_{\mathbf{x}} f(\mathbf{x}) log \frac{f(\mathbf{x})}{g(\mathbf{x})}$$

The higher KLD between two different sources solutions at the same antenna, the solver noise is less

- We fit a Gaussian mixture model to the solutions by EM algorithm
- KLD is calculated via Monte-Carlo algorithm

Introduction

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Calibration

- Estimating unknown instrument and the sky parameters and correcting them before imaging.
- Finding the Maximum Likelihood estimate of the sky and the Instrument Unknown parameters.

Calibration challenges:

- The scientific goals of LOFAR require extreme sensitivity and dynamic range (LOFAR EOR)
- Intrinsically polarized feeds (dipoles) - polarized measurement equation
- Wide fields of view
- Pronounced direction-dependent effects

Input noise "instrumental noise" → Calibration Algorithm → output noise Noise in calibration solutions

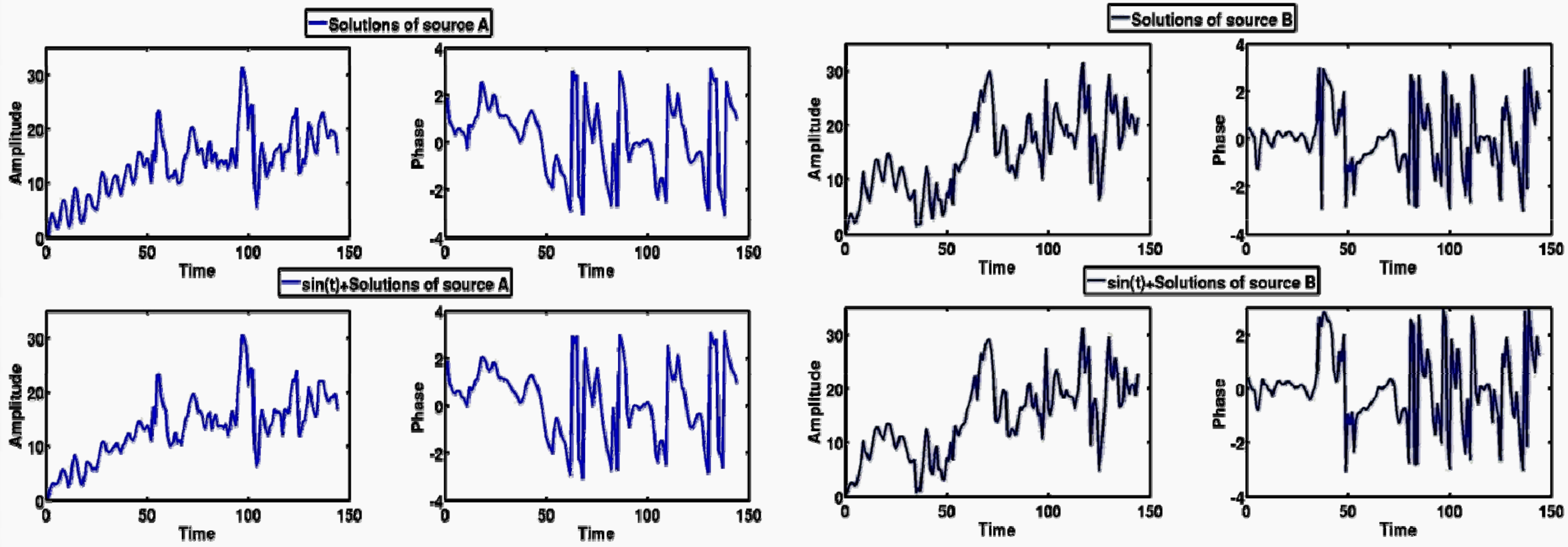
Solver noise

Illustrative example

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Detection solver noise via KLD and LRT

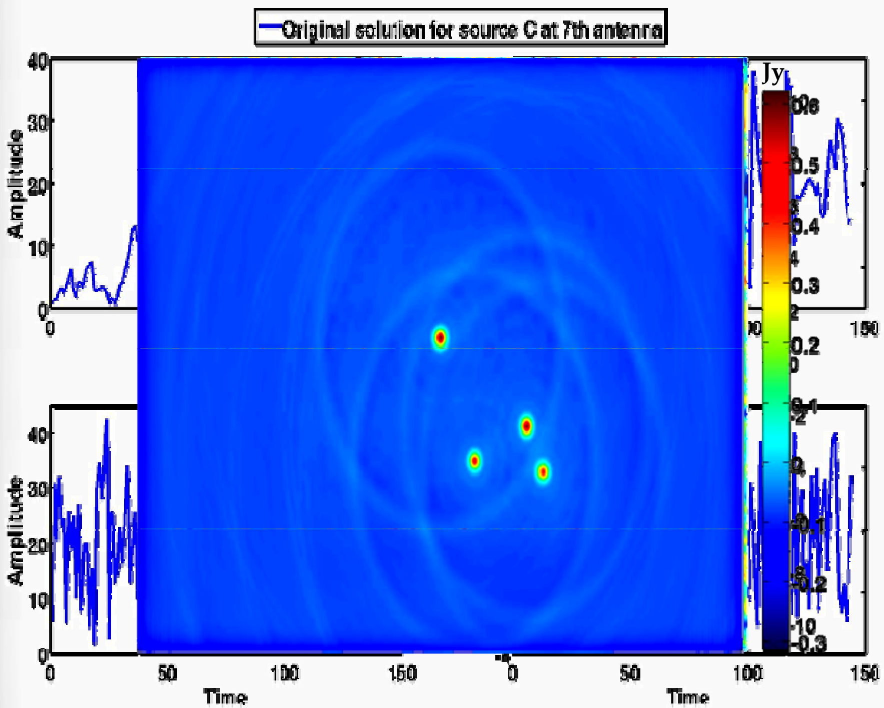
Solutions of antenna seven are added *sin(t)* noise where *t* is integration time:



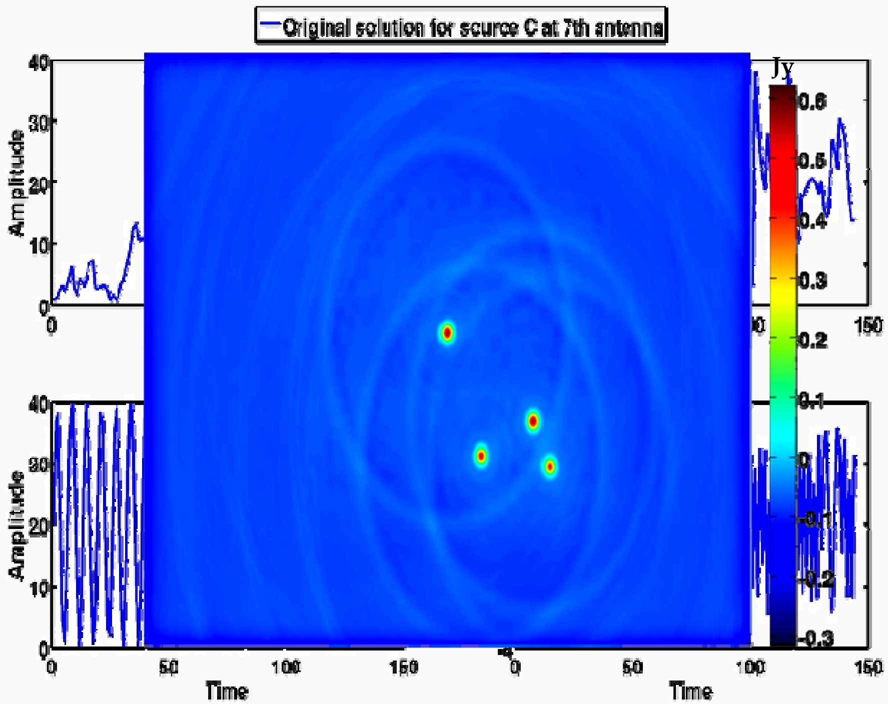
Illustrative example

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html

Direction dependent gains

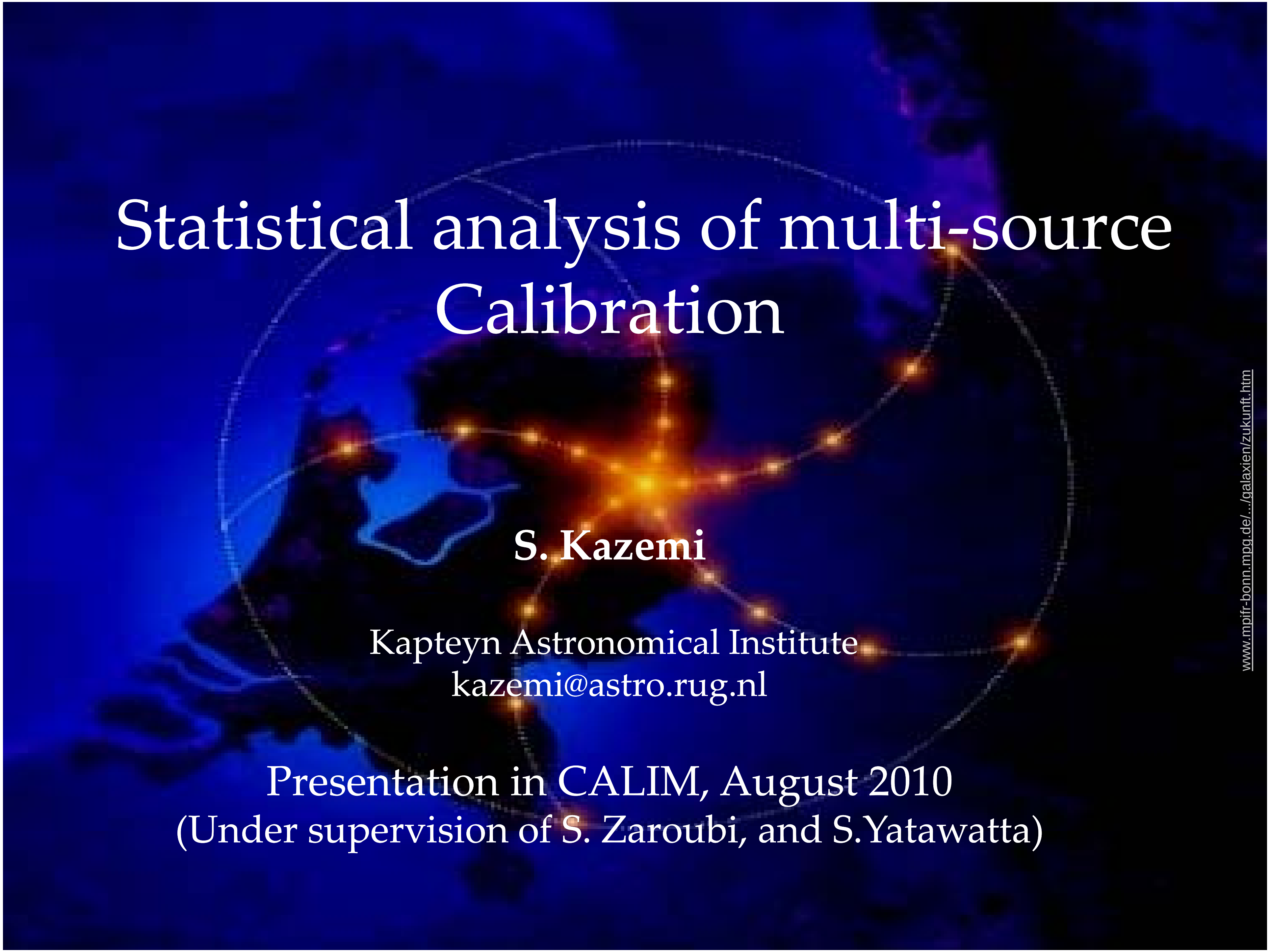


Random solutions



Sin(t) solutions

Xq,s



Statistical analysis of multi-source Calibration

S. Kazemi

Kapteyn Astronomical Institute
kazemi@astro.rug.nl

Presentation in CALIM, August 2010
(Under supervision of S. Zaroubi, and S.Yatawatta)

Illustrative example

rai.ncsa.uiuc.edu/RAI_Projects_SKA.html