Jacobian Leverage as a Diagnostic in Radio Interferometric Calibration

Sarod Yatawatta

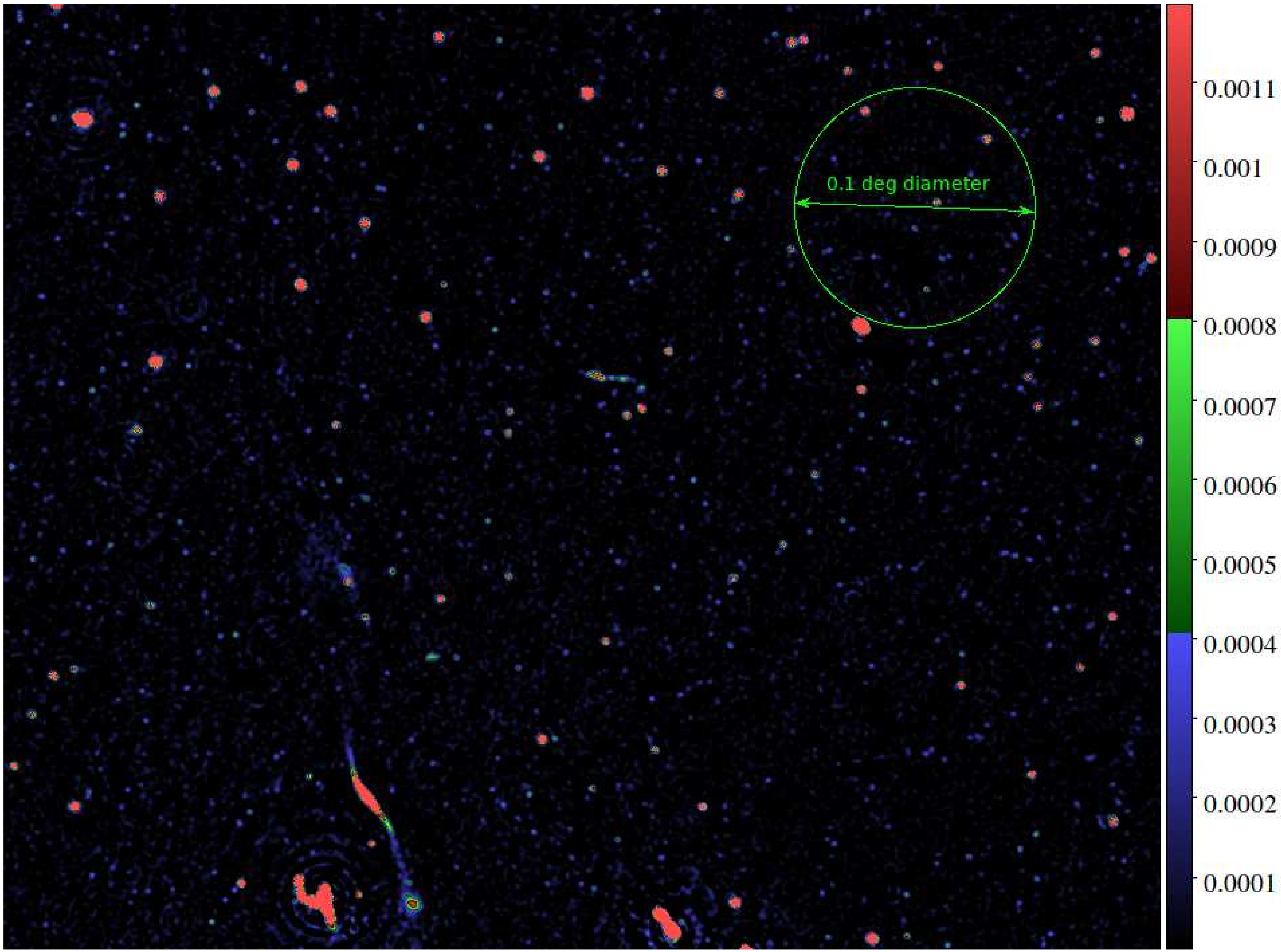
ASTRON

The Netherlands

Introduction

- ☐ Radio telescopes are getting bigger: WSRT: 14 stations, LOFAR: 72 stations, SKA-Low: 512 stations.
- ☐ SKA: Higher sensitivity, more data, more stations.
- ☐ Calibration: essential for correcting systematic errors (beam,ionosphere), removal of foregrounds (Epoch of Reionization).
- ☐ Big data in radio astronomy: 100s of thousands of unknowns, millions of constraints: how can we make sure calibration works as expected?

LOFAR Deep Image



150 MHz, 2$''$ pixels, 40 μJy noise, dynamic range $>$ 150 000

Calibration

uncalibrated image

what we want

what we don't want

Diagnostics

- ☐ Looking at final image: caveat : a nice clean image can be misleading.
- ☐ Tools from estimation theory : Cramer Rao lower bound.
- ☐ Tools from statistics : cross validation and leverage.

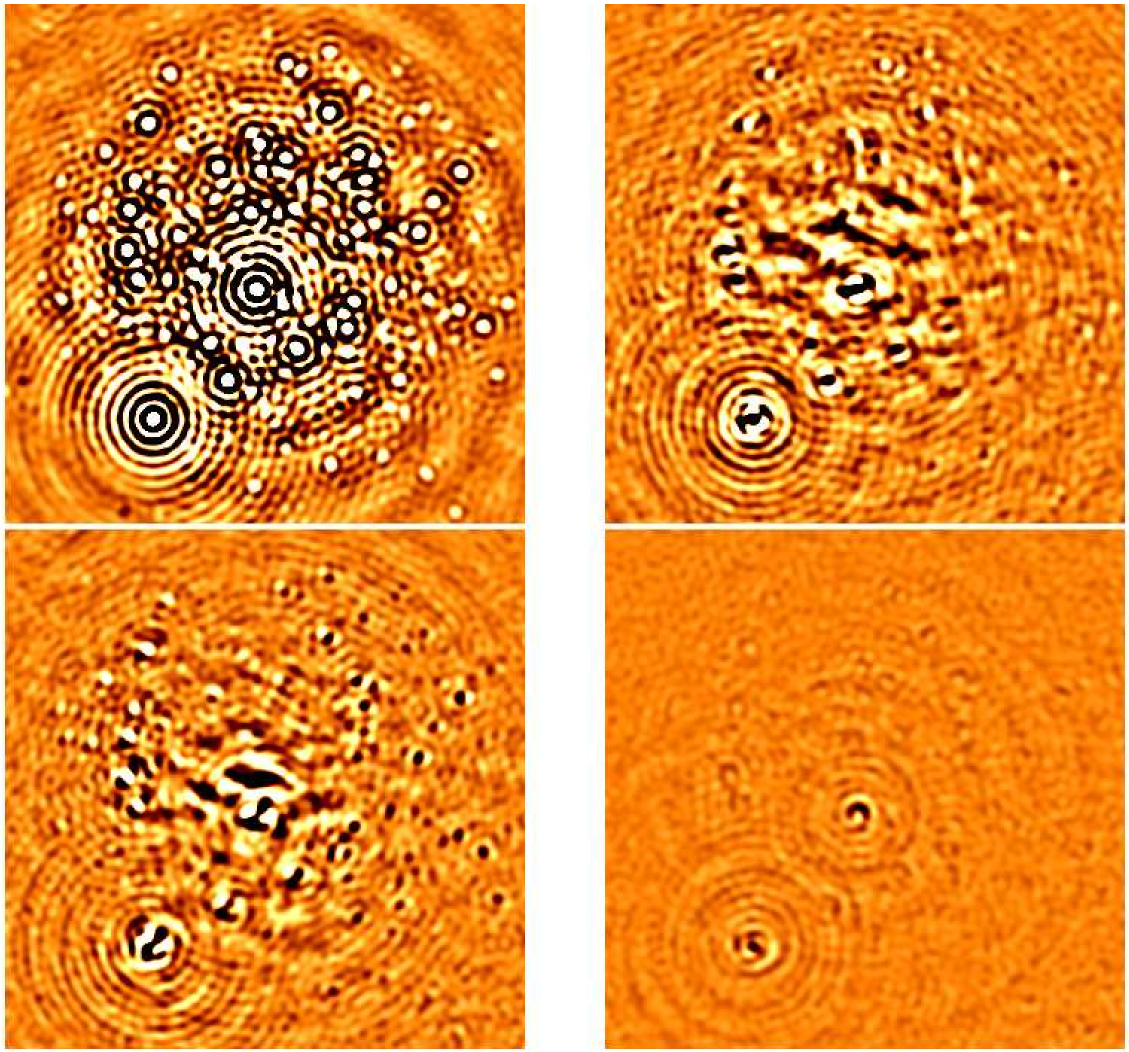
$$\underbrace{\mathbf{y}}_{\text{data}} = \underbrace{\mathbf{m}\left(\underbrace{\boldsymbol{\theta}}_{\text{parameters}}\right)}_{\text{model}} + \underbrace{\mathbf{n}}_{\text{noise}}$$

Calibration: obtain $\widehat{\boldsymbol{\theta}}$ by fitting $\mathbf{m}(\boldsymbol{\theta})$ to $\mathbf{y}$. Cramer Rao lower bound bounds variance of $\widehat{\boldsymbol{\theta}}$.
But most science is in the residual

$$\mathbf{r} = \mathbf{y} - \mathbf{m}(\widehat{\boldsymbol{\theta}})$$

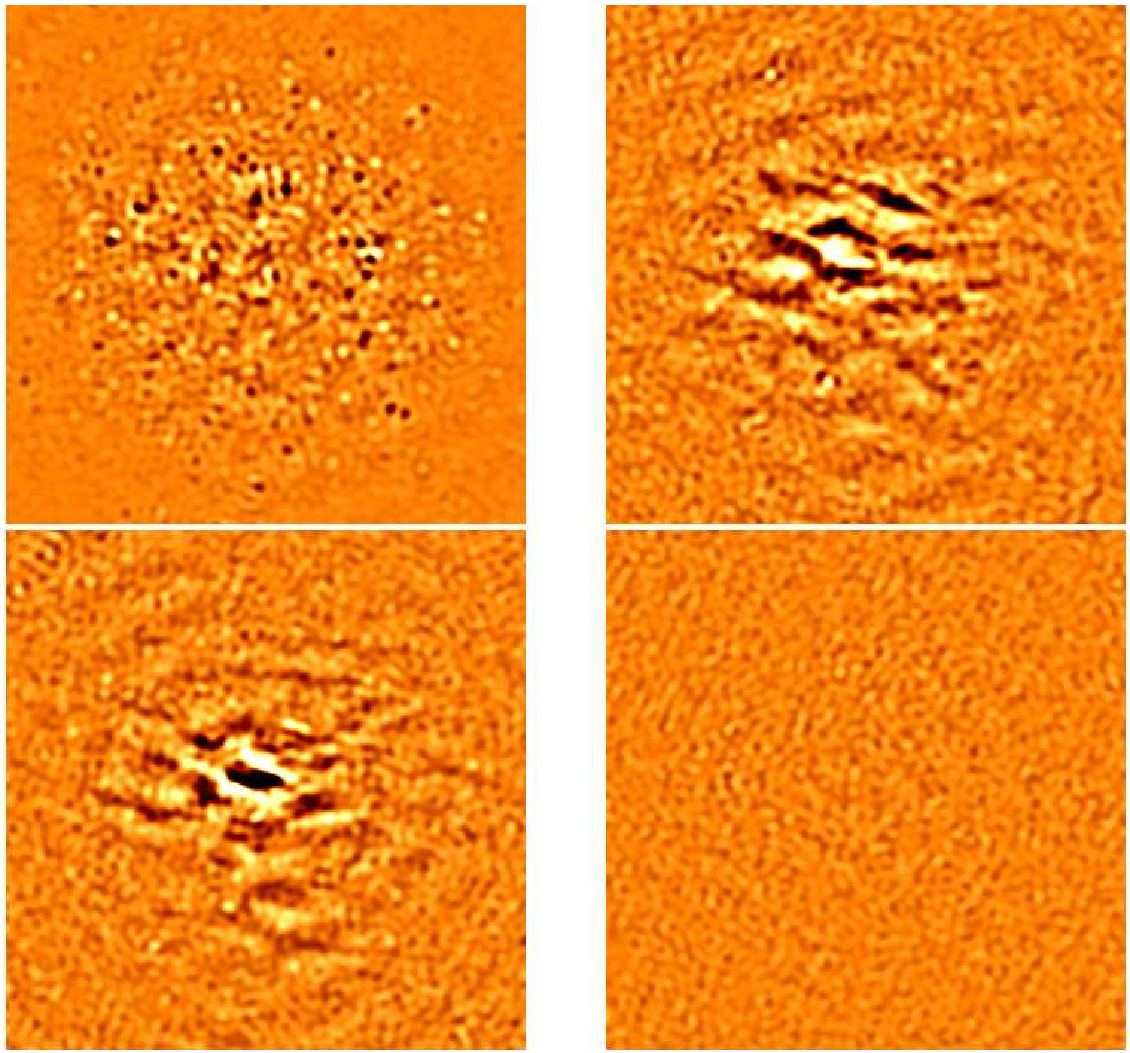
Relating $\mathrm{Var}(\widehat{\boldsymbol{\theta}})$ to $\mathrm{Var}(\mathbf{r})$ is not easy.
Cross validation: using only part of the data to calibrate and use the excluded part for validation.

Image before calibration



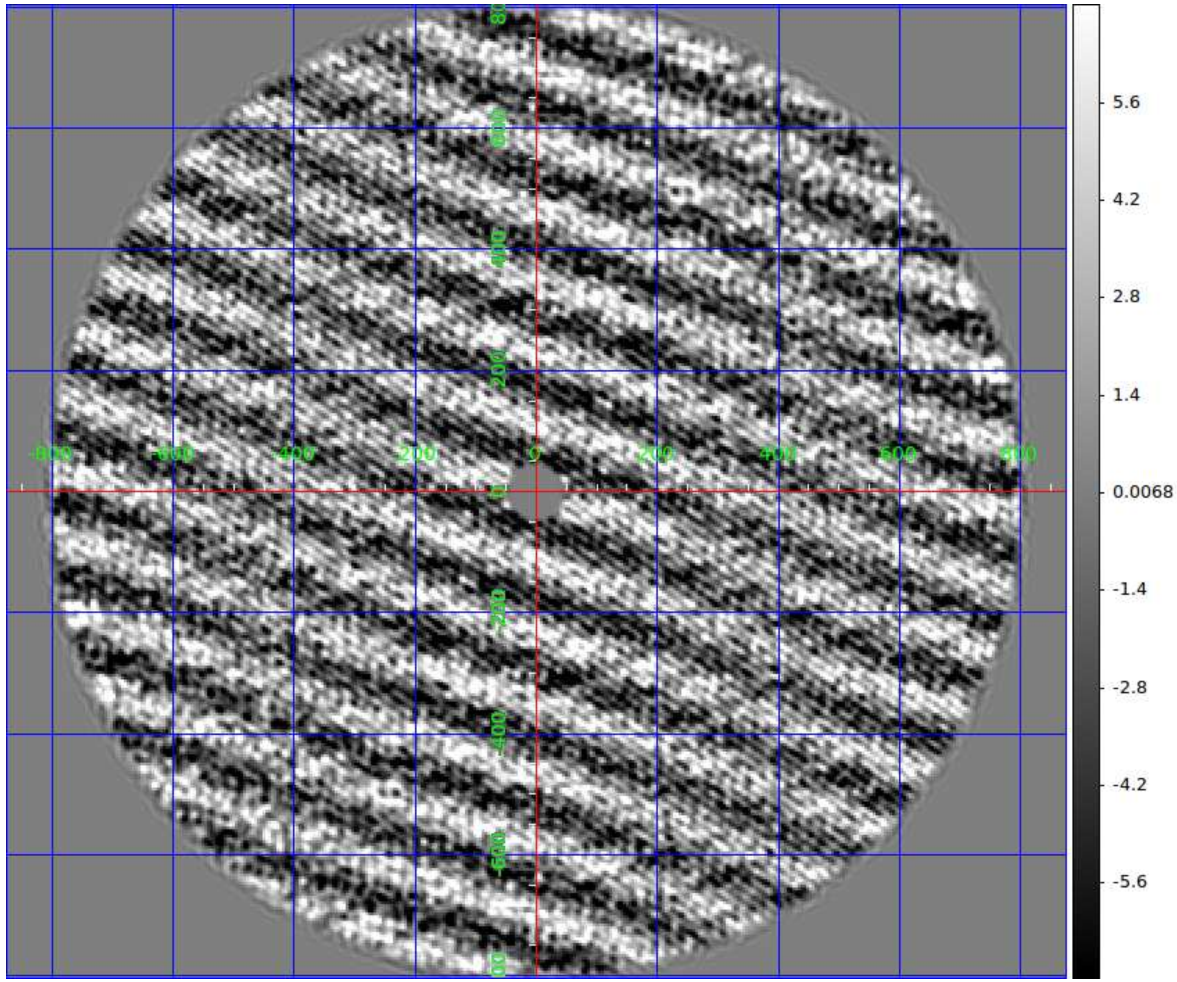
I,Q,U,V images baselines $<=$ 250 wavelengths

Image after calibration



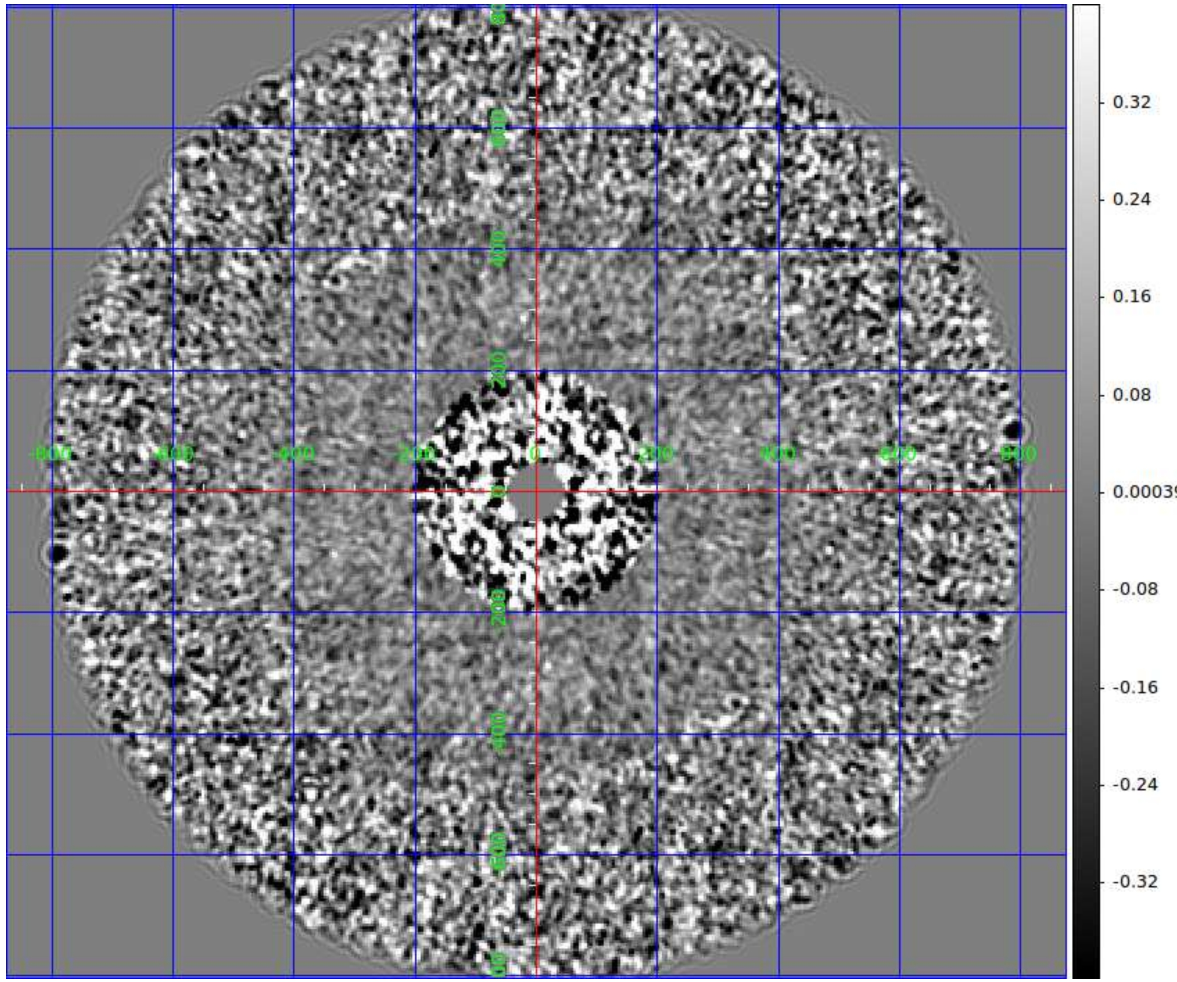
I,Q,U,V calibration using baselines $>$ 250 wavelengths

Gridded data before calibration



Real part of I, baselines $<=$ 800 wavelengths

Gridded data after calibration



calibration using baselines $>$ 250 wavelengths

Linear Example

Linear system with tall, full rank matrix $\mathbf{A}$, zero mean, white Gaussian noise $\mathbf{n}$

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n} \Rightarrow \widehat{\mathbf{x}} = \mathbf{A}^{\dagger}\mathbf{y}$$

Least squares solution using full data: $\widehat{\mathbf{x}}$

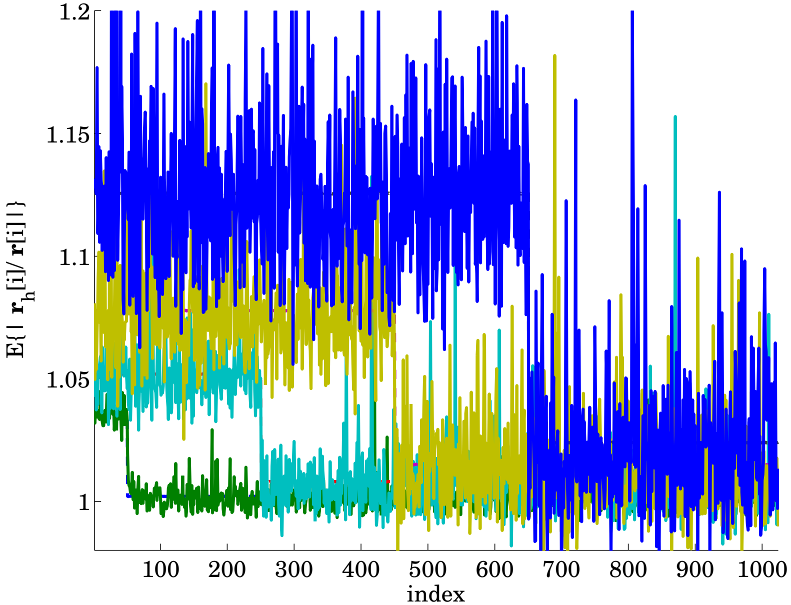
$$\begin{bmatrix} \mathbf{y}_l \\ \mathbf{y}_h \end{bmatrix} = \begin{bmatrix} \mathbf{A}_l \\ \mathbf{A}_h \end{bmatrix} \mathbf{x} + \mathbf{n} \Rightarrow \widehat{\mathbf{x}_h} = \mathbf{A}_h^{\dagger}\mathbf{y}_h$$

Least squares solution using a subset of data $\mathbf{y}_h$: $\widehat{\mathbf{x}_h}$.
Residual calculated using full data $\mathbf{r}$ and a subset of data $\mathbf{r}_h$

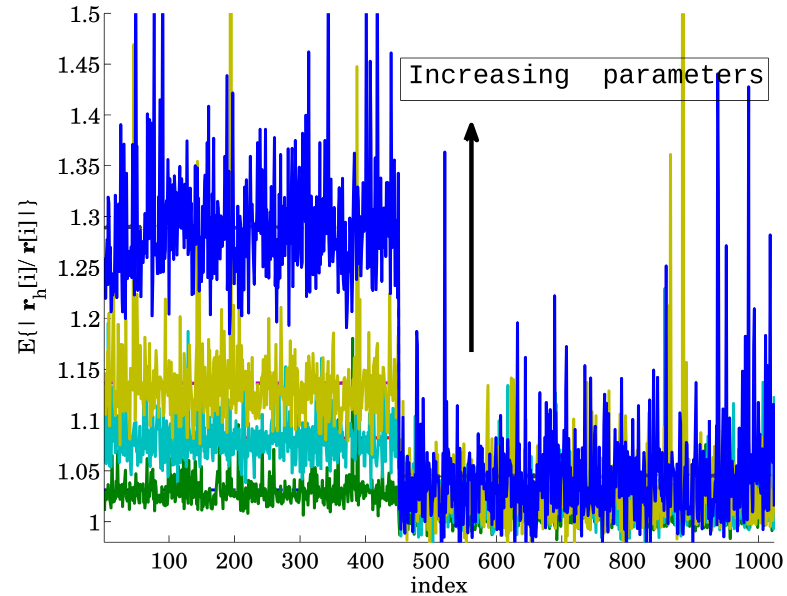
$$\mathbf{r} = \mathbf{y} - \mathbf{A}\widehat{\mathbf{x}} \qquad \mathbf{r}_h = \mathbf{y} - \mathbf{A}\widehat{\mathbf{x}_h}$$

How are $\mathbf{r}$ and $\mathbf{r}_h$ related? Plot $E\{|\mathbf{r}_h[i]/\mathbf{r}[i]|\}$ for random realizations of $\mathbf{A}$ and $\mathbf{n}$.

Residuals



50, 250, 450 and 650 datapoints excluded out of 1024



450 datapoints excluded out of 1024

- Variance of residuals increase as more data are excluded.
- Jump of excluded residual increases as more data are excluded.

Complete Model

To model data exclusion, introduce artificial variables γ (size equal to excluded data points).

$$\begin{bmatrix} \mathbf{y}_l \\ \mathbf{y}_h \end{bmatrix} = \begin{bmatrix} \mathbf{A}_l \\ \mathbf{A}_h \end{bmatrix} \mathbf{x} + \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \boldsymbol{\gamma} + \mathbf{n}$$

What we really solve is

$$\begin{bmatrix} \mathbf{y}_l \\ \mathbf{y}_h \end{bmatrix} = \begin{bmatrix} \mathbf{A}_l & \mathbf{I} \\ \mathbf{A}_h & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{x} \\ \boldsymbol{\gamma} \end{bmatrix} + \mathbf{n}$$

with unknowns $\mathbf{x}$ and $\boldsymbol{\gamma}$. Cramer-Rao lower bound gives bounds for variance of $\widehat{\mathbf{x}}$ and $\widehat{\boldsymbol{\gamma}}$. ($E\{\widehat{\mathbf{x}}\} = \mathbf{x}^\star$, $E\{\widehat{\boldsymbol{\gamma}}\} = \mathbf{0}$).
Variance of residuals: of data included in solution

$$\mathrm{Var}\left(\mathbf{r}_h\right) \propto \mathrm{Var}\left(\widehat{\mathbf{x}}\right)$$

of data excluded from solution

$$\mathrm{Var}\left(\mathbf{r}_l\right) \propto \mathrm{Var}\left(\widehat{\mathbf{x}}\right) + \mathrm{Var}\left(\widehat{\boldsymbol{\gamma}}\right)$$

Both can be calculated in closed form for a linear system.

Leverage

Data $\mathbf{y}$ gives $\widehat{\boldsymbol{\theta}}$ and data $\mathbf{y}+b\boldsymbol{f}$ gives $\widehat{\boldsymbol{\theta}}_b$ as estimates.
Leverage vector

$$\boldsymbol{g} \stackrel{\triangle}{=} \lim_{b\to 0}\frac{1}{b}\left(\mathbf{m}(\widehat{\boldsymbol{\theta}}_b)-\mathbf{m}(\widehat{\boldsymbol{\theta}})\right)$$

[St. Laurent & Cook, 1992] Jacobian leverage $\boldsymbol{\Gamma}(\boldsymbol{\theta})$ directly gives

$$\boldsymbol{g}=\boldsymbol{\Gamma}(\widehat{\boldsymbol{\theta}})\boldsymbol{f}$$

This can be used to directly find $\mathrm{Var}(\mathbf{r})$. For calibration, datapoints included have leverage $\Gamma^{ii}(\widehat{\boldsymbol{\theta}})$ and datapoints excluded have leverage $\Gamma^{ii}(\widehat{\boldsymbol{\theta}})+1$. Variance of residuals [Patil et al., in prep.] of data included in calibration

$$\mathrm{Var}\left(\mathbf{r}_h\right)\propto\left(\Gamma^{ii}(\widehat{\boldsymbol{\theta}})\right)^2\mathrm{Var}\left(\widehat{\boldsymbol{\theta}}\right)$$

of data excluded from calibration

$$\mathrm{Var}\left(\mathbf{r}_l\right)\propto\left(\Gamma^{ii}(\widehat{\boldsymbol{\theta}})+1\right)^2\mathrm{Var}\left(\widehat{\boldsymbol{\theta}}\right)$$

Conclusions

- ☐ Cross-validation and Leverage: directly estimate variance of residuals.
- ☐ Excluding baselines in calibration (cross validation) : Diffuse structure, Calibration transfer, model incompleteness : will increase noise.
- ☐ Preserving weak signals in residuals: use complete sky model, use noise model with heavy tails (Student's T, Huber), iterative weighting of data (pre-whitening) [Kazemi, Yatawatta, 2013].
- ☐ Increasing constraints : distributed calibration [Yatawatta, 2015] ($=$ adding regularization to calibration).

Acknowledgments: European Research Council Advanced Grant LOFARCORE - 339743.