Astronomical Data Analysis Software and Systems XIII
ASP Conference Series, Vol. 314, 2004
F. Ochsenbein, M. Allen, and D. Egret, eds.

The LOFAR Global Sky Model: Some Design Challenges

Oleg M. Smirnov, Jan E. Noordam

Netherlands Foundation for Research in Astronomy (ASTRON)

Abstract.
The LOFAR Global Sky Model (GSM) will be an all-sky database of some 100 million objects, with flux & polarization measurements in the 20–200 MHz range. The primary function of the GSM is to support LOFAR calibration and data reduction. The GSM is expected to provide a model of all sufficiently bright sources in any given field, having enough detail and precision to calibrate and subtract these sources and yield residual images of the faint background. The GSM is expected to be continuously updated and refined during LOFAR operation in a "closed loop" of sorts. The GSM is also a valuable stand-alone data product that can be made compatible with the VO.

The instrumental characteristics of LOFAR pose large challenges to GSM design, some of them unique even in the field of very large catalogs. Besides sheer size, this includes highly complex source models (thus making for a very non-uniform database structure), stringent performance requirements for operational use, the need to update source models operationally, and various data management issues. This paper will focus on some of these challenges, discuss our approaches to dealing with them, and present a prototype GSM being developed for the LOFAR Pilot Selfcal System (PSS).

1. Introduction

LOFAR[1] is a a large, distributed radio telescope being designed by an international consortium (ASTRON, ATNF, MIT Haystack, NRL). Some architectural features of the current design are:

- 20–200 MHz range;
- > 100 phased array stations combined into an aperture synthesis array;
- Log-spiral configuration with a dense central core, largest baseline is ≈300 km.

The unique instrumental characteristics of LOFAR pose new challenges to calibration. This has led us to formulate a requirement for a Global Sky Model (GSM):

[1] `http://www.lofar.org`

- Calibration of LOFAR for any observation requires reference sources from all over the sky, due to the extensive side lobes of the instrument.
- The GSM itself, being a very large source catalog updated over the lifetime of the telescope, is one of the main LOFAR deliverables. As such, it will be made available within the Virtual Observatory.

2. GSM: the next step in radio sky models

The self-calibration algorithm employed in radio astronomy is simple in essence, but extremely difficult in the details. The basic selfcal iteration step consists of predicting the sky, applying instrumental effects, and comparing the results with observed data. The current record for dynamic range with selfcal is 10^6, achieved at the WSRT (de Bruin 1996). The LOFAR design target is 10^8. A proof of concept study will be carried out using WSRT data, which can potentially yield 10^7, given a sophisticated enough calibration approach.

One of the main limiting factors in the dynamic range of selfcal is accuracy of the predict step. An accurate predict requires a sufficiently complex and flexible flux source representation, and a complex enough *Measurement Equation* (Hamaker, Bregman & Sault 1996).

- The traditional sky model – still widely used today by packages such as AIPS[2] – is a collection of CLEAN components. This approach can handle extended sources with some success, but fails to take advantage of physical continuity, and fails to represent variability in time. Various refinements of CLEAN have been developed over the years to address these shortcomings.
- NEWSTAR[3], the WSRT data reduction package, introduced a parameterized source representation ($IQUV$ fluxes, rotation measure, spectral index), as well as elliptical gaussians to model extended sources. CLEAN components are still used in NEWSTAR to represent low-level extended sources.

LOFAR (and other future instruments such as the Square Kilometer Array) requires us to take the next step. This includes more sophisticated source models, in particular spatially extended sources, as well as time and frequency variability of all source parameters.

2.1. Some features of the LOFAR GSM

- Expected size: 100 million objects. All-sky coverage.
- Fluxes and polarizations in the 20–200 MHz range.
- Very rich with spatially extended sources. Needs sophisticated parameterizations, with all parameters potentially being functions of time and frequency.
- Must support a variety of object morphologies: point sources, extended sources as shapelets, pixons, CLEAN components, images, etc. Morphology can change qualitatively with frequency.

---

[2] `http://www.aoc.nrao.edu/aips/`

[3] `http://www.astron.nl/newstar/`

- Initially populated from existing catalogs and updated over the lifetime of the instrument.
- One of the two main deliverables of LOFAR.

2.2. The GSM life cycle

Operationally, the GSM software must support the following life cycle:

1. For any observation, a subset is extracted into a Local Sky Model (LSM);
2. Calibration is performed, the LSM is updated with better parameters for existing sources and possibly new sources;
3. Changes are optionally committed back to the GSM;
4. After imaging (which produces the data products), images of faint sources may also be added to the GSM.

3. Data representation in the GSM

The question of data representation has emerged as one of the central issues of GSM design. This representation must be flexible enough to capture sufficient detail for calibration and imaging. Because of the GSM–calibration–GSM "closed loop", the notion of parametrized sources must be explicitly supported by the GSM itself. Thus, what is really required is a non-uniform and extensible source representation.

3.1. MeqTrees

MeqTrees (Measurement Equation Trees) are a mechanism being developed in the LOFAR Pilot Selfcal System (PSS). A MeqTree[4] corresponds to a mathematical expression. The leaf nodes represent parameters and data sources, while nodes down the tree represent derived expressions, such as the predicted visibility of a source in a given direction.

A MeqTree can recursively evaluate its expression, and estimate partial derivatives w.r.t. specific parameters. Thus, a MeqTree can be employed to iteratively solve an equation for a set of parameters. All parameters in the tree (*MeqParms*) are polynomials (and in the future, possibly other smooth functions) of frequency and time, so the actual solvables are the individual polynomial coefficients.

Since any mathematical equation can be represented in tree form, PSS-4 (the current development cycle of PSS) should be able to solve for arbitrary measurement equations. Operationally, trees are constructed from a scripting layer, while a fast C++ kernel to evaluate and solve them. In PSS-4, the scripting is provided by the Glish language of AIPS++[5]; support for other languages (primarily Python) will be added in future cycles.

[4]Strictly speaking, these are not simply trees, but rather a more general type of graph – the directed acyclic graph (DAG); the name *MeqTree* is now entrenched for historical reasons.

[5]`http://aips2.nrao.edu/docs/aips++.html`

3.2. The GSM as a collection of trees

MeqTrees provide a perfect answer to the issue of GSM data representation. To predict the sky, we need to be able to predict the visibility contribution of each source for a given baseline and pointing direction. This, however, may need to be computed differently for different types of sources. By representing a source (or rather, the components of a source, such as the Stokes I flux) as a MeqTree, we reap a number of benefits:

- **Polymorphism:** all sources appear the same way (that is, as the root node of of the specific source component sub-tree) both to the user and to the calibration system, regardless of the details of internal representation.
- **Extensibility:** source structure can be represented and parametrized to any level of complexity.
- **Parallelization:** trees allow for fine-grained parallelism, which is very important considering the immense computing requirements of LOFAR calibration.
- **Rapid experimentation:** trees can be torn down and rebuilt at run-time. This allows for quick experiments with alternative source representations.

Of course, this approach also incurs a number of trade-offs:

- **Computational overhead:** trees require more complicated housekeeping and data management compared to a "hard-wired" source representation.
- **DB complications:** this type of structure is rather difficult to represent in a traditional relational database.

4. Development plan

An initial version of the LOFAR GSM (GSM-0), is being developed as part of the PSS-4 cycle. This is meant as proof-of-concept implementation, thus we have set the following modest goals:

- A homogeneous source representation borrowed from NEWSTAR, with a fixed set of source parameters: RA/Dec, $IQUV$ fluxes at a reference frequency, spectral index, rotation measure. Extended sources will be modeled by elliptical gaussians requiring three additional parameters: major/minor axis size, and orientation.
- All parameters can be 2D polynomials of time and frequency.
- GSM-0 is stored in a simple flat table. In the background, we will work toward a full database implementation of MeqTrees.
- Internally, PSS-4 will use fully-featured MeqTrees. These will be constructed on-the-fly in the GSM-0 code, using parameters from the GSM table. Thus, we will be able to test calibration will fully featured MeqTrees, but the structure of the trees will be hardwired in GSM-0.

References

de Bruin, G. 1996, in High-Sensitivity Radio Astronomy, ed. D. Jackson, 233

Hamaker, J.P., Bregman, J.D., Sault, R.J. 1996, A&AS, 117, 137