SPAM: A data reduction recipe for high-resolution, low-frequency radio-interferometric observations

H. T. Intema[1] *

[1]*National Radio Astronomy Observatory, 1003 Lopezville Road, Socorro, NM 87801-0387, USA*

Abstract. High-resolution astronomical imaging at sub-GHz radio frequencies has been available for more than 15 years, with the VLA at 74 and 330 MHz, and the GMRT at 150, 240, 330 and 610 MHz. Recent developments include wide-bandwidth upgrades for VLA and GMRT, and commissioning of the aperture-array-based, multibeam telescope LOFAR. A common feature of these telescopes is the necessity to deconvolve the very many detectable sources within their wide fields-of-view and beyond. This is complicated by gain variations in the radio signal path that depend on viewing direction. One such example is phase errors due to the ionosphere.

Here I discuss the inner workings of SPAM, a set of AIPS-based data reduction scripts in Python that includes direction-dependent calibration and imaging. Since its first version in 2008, SPAM has been applied to many GMRT data sets at various frequencies. Many valuable lessons were learned, and translated into various SPAM software modifications. Nowadays, semi-automated SPAM data reduction recipes can be applied to almost any GMRT data set, yielding good quality continuum images comparable with (or often better than) hand-reduced results. SPAM is currently being migrated from AIPS to CASA with an extension to handle wide bandwidths. This is aimed at providing users of the VLA low-band system and the upcoming wide-bandwidth GMRT with the necessary data reduction tools.

Keywords : atmospheric effects – methods: data analysis – instrumentation: interferometers

---

*email: hintema@nrao.edu

1.   Introduction

Low-frequency radio telescopes have relatively large fields-of-view, because the telescope primary beam size scales with wavelength. Current high-resolution (sub-arcminute) radio interferometers that operate at sub-GHz frequencies, like the Giant Metrewave Radio Telescope (GMRT), the LOw-Frequency ARray (LOFAR), or the Very Large Array (VLA), have field diameters measured in (sometimes tens of) degrees. Typical challenges of handling high-resolution, low-frequency continuum observations are (i) imaging and deconvolving many sources inside (and outside!) the large primary beam area, (ii) direction-dependent (DD) visibility amplitude and phase variations across the field-of-view due to antenna beam patterns, pointing errors, ionosphere, etc., and the abundance of Radio Frequency Interference (RFI).

Processing data coming from these telescopes can be daunting, since all the challenges mentioned above (and more) needs to be dealt with at the same time. In data reduction, the *art* is to determine at any given time which effect is dominant in limiting the image quality. Generally, a few *big effects* stand out and are generally easy to identify and fix. But subsequenty, many *smaller effect* will present themelves, which may be more difficult to recognize, and harder to identify and disentangle. A effective way to mitigate the smaller effects is to perform iterations of (DD) calibration, imaging and flagging, in each step refining the quality of both the visibility data and the reconstructed sky model.

2.   SPAM: an overview

Developments towards a semi-automated data reduction package started in 2006, driven by the complicated reduction of GMRT 150 MHz observations on galaxy cluster Abell 2256 (and other similar data sets). Several challenges presented themselves at the same time: (i) the target contains faint, diffuse emission in the presence of bright interfering sources, (ii) the data was strongly affected by RFI on many baselines, (iii) the data was strongly affected by ionospheric phase distortions, (iv) the data was affected by instrumental instabilities. It quickly became clear that this data required an algorithm to correct for ionospheric phase errors, and possibly other DD effects. Simultaneously followed the desire to automate several trivial but very time-consuming data reduction steps.

For the basis of these new developments a stable data reduction package was sought, with the possibility to easily expand the functionality. This was found in the combination of the data reduction package AIPS (Greisen 2003) and the powerful high-level programming language Python (and its standard scientific libraries like scipy, pylab, matplotlib, and numpy), with the ParselTongue interface (Kettenis et al. 2006) providing access to AIPS tasks, data files (images and visibilities) and tables from Python. A new Python module was created named *SPAM*, an acronym for *Source Peeling and Atmospheric Modeling*, encapsulating high-level data reduction functions and new algorithm development.

SPAM-based data reductions, i.e. for GMRT observations, have been captured in Python

scripts that execute AIPS tasks directly (mostly during initial reduction steps), call high-level functions that encompass multiple AIPS /ParselTongue calls, and require few manual operations. These scripted data reductions automatically keep a history of the data reduction steps, and provide a standard that is well-tested and reproducible. At the same time, the SPAM user is required to perform some manual steps, and can easily decide to change the order of data reduction steps depending on the target field geometry and the intended science, therefore the term data reduction *recipe* is more appropriate than *pipeline*. Note that SPAM assumes an unpolarized sky, observed with limited ($\lesssim 20$ percent) fractional bandwidth.

SPAM recipes are available for all sub-GHz GMRT frequencies, developed and refined over the last several years, with minimal differences between frequencies. The general structure of this data reduction is described in Intema et al. (2009), and includes *measuring* DD ionospheric phase errors through peeling of bright sources within the science field (e.g., Noordam 2004), *modeling* these errors with a single- or multi-layer ionospheric phase model, and *applying* ionospheric phase corrections during wide-field (re-)imaging of the science field (see Section 2.1).

A combination of RFI- and bad data mitigation routines are used inbetween rounds of calibration and imaging. *Classical outlier removal* excises visibilities mostly based on excessive visibility amplitudes and statistical outlier rejection along the time and frequency axes. RFI *subtraction* models and subtracts low-level, quasi-continuous, ground-based RFI based on its fringe-rotation signature (e.g., Athreya 2009). And *ripple killing* excises bad visibility data based on a combination of high visibility amplitudes, high visibility weights and high imaging weights (i.e., though uniform weighting). This is done by Fourier-transforming residual images back onto the UV-plane, identifying high-amplitude UV-cells, and rejecting all visibilities that fall in these cells[1]. A more elaborate description will be given in an upcoming paper (Intema et al., in preparation).

2.1 SPAM ionospheric calibration and imaging

Following Lonsdale (2005) there are four ionospheric calibration regimes, depending on the size of the radio interferometer array and the size of the field-of-view, both relative to the scale of phase structure in the ionosphere. This makes most sense when imagining a single horizontal density wave (or a bubble) in a thin layer at a fixed height. Regime 4 is the most complex case, when both the horizontal size of the array and the projected field-of-view at ionospheric height are similar or larger than the horizontal scale of the ionospheric phase structure. In this regime, when looking towards a single source in the field-of-view, the higher-order phase structure over the array will cause both an apparent position shift and a source deformation. When looking at a second source in the same field-of-view at sufficient distance from the first, both the apparent position shift and the source distortion will be different.

---

[1]E.g., see the AIPS tutorial by Enßlin & Kronberg
`http://www.mpa-garching.mpg.de/~ensslin/Paper/Reduce/remove.ps.gz`

The SPAM ionospheric calibration strategy is designed to operate in regime 4, meaning that it will also work in the less complex regimes 1–3. In an automated peeling routine, measurements of the ionospheric phase structure are obtained by phase calibrating on (typically 10–20) bright sources within the field-of-view. Per time interval (typically 10–20 seconds), the measured phases of all source–antenna pairs are mapped onto a common domain along their line-of-sight, i.e., at the pierce points through a virtual phase screen at fixed height. The phases are fit with an optimized set of base functions, reproducing the measured phases, and predicting DD phases corrections in arbitrary viewing directions during imaging.

The automated wide-field imager applies any relevant DD phase corrections on the fly in a modified version of polyhedron (facet-based) imaging and Cotton-Schwab CLEAN deconvolution (e.g., see Taylor, Carilli & Perley 1999). The intrinsic book-keeping issues of faceted imaging are largely hidden from the user. The imager also includes an iterative, robust clean boxing scheme, and simultaneous imaging of the primary beam area and bright outlier fields. In principle, the generic imager can handle other DD effects as well, like asymmetric primary beams or pointing errors, as long as a model is available to generate the DD correction tables.

3. Summary

SPAM is an AIPS-based Python package that provides semi-automated data reduction scripts for all sub-GHz frequencies at GMRT. Apart from well-tested standard data reduction steps, SPAM includes direction-dependent (ionospheric) calibration and imaging, and alternative RFI- and bad data mitigation methods. With SPAM, GMRT data reductions are highly efficient, highly reproducible, and give good quality results. SPAM is available for everyone to use[2].

SPAM for CASA, currently under development, will also provide wide-bandwidth ionospheric calibration, direction-dependent ionospheric Faraday rotation calibration, and will naturally benefit from the advanced modes (and continuous improvements) of the CASA imager (e.g., see Bhatnagar et al. 2013, and references therein).

References

Athreya R., 2009, ApJ, 696, 885
Bhatnagar S., Rau U., Golap K., 2013, ApJ, 770, 91
Greisen E. W., 2003, ASSL, 285, 109
Intema H. T., van der Tol S., Cotton W. D., et al., 2009, A&A, 501, 1185
Kettenis M., van Langevelde H. J., Reynolds C., Cotton B., 2006, ASPC, 351, 497
Lonsdale C. J., 2005, ASPC, 345, 399
Noordam J. E., 2004, SPIE, 5489, 817
Taylor G. B., Carilli C. L., Perley R. A., 1999, ASPC, 180

---

[2]`https://safe.nrao.edu/wiki/bin/view/Main/HuibIntemaSpam`