AIPS Memo 112

Capabilities of the VLA pipeline in AIPS

Lorant O. Sjouwerman

March 19, 2007

Abstract

This document describes the VLA pipeline procedure. The procedure runs in AIPS, though a system has been set up to process VLA data with this pipeline from a UNIX command line. The latter and an analysis of a pilot project on 1999/2000 VLA B-array 5 and 8 GHz continuum data run using this pipeline are described in another document; most of this document is to explain about the capabilities of the pipeline in combination with AIPS input parameters, choices and limitations, as invoked from inside AIPS. Starting conditions and suggestions for success are given. This document covers the status of the VLA pipeline as per mid-November 2006.

Contents

1 Introduction to the early versions **2**

2 The VLA pipeline setup **2**

3 The VLA pipeline in AIPS **3**

3.1 VLARUN calibration input parameters . . . 3

3.2 VLARUN imaging input parameters . . . 4

3.3 VLARUN self-cal input parameters . . . 5

3.4 VLARUN pipeline parameters . . . 5

4 Running the VLA pipeline **6**

4.1 Starting conditions . . . 6

4.2 Suggested use . . . 6

5 Common problems and fixes **7**

6 Improvements **7**

1 Introduction to the early versions

An early version (the "first" version) of the VLA pipeline was simple and created for private use. It was intended to perform calibration and imaging of high-frequency multi-IF spectral line VLA survey data in any array configuration using NRAO's classic AIPS package. It assumed that all bad data was removed and that one of the standard VLA flux calibrators was included in the observations. A cleaned-up version of this was released to the general public as an AIPS run-file procedure named "VLARUN" in 2003, and used for demonstrations at the Santa Fe "xraydio" meeting in February 2004. This pipeline was a very useful tool to examine the u, v-data and create images within minutes of acquiring the raw VLA data, point out the trouble spots (i.e. bad data), and obtain first order self-cal source models.

Since then the VLA pipeline has been further developed for automated pipelining, with emphasis on X- and C-band continuum in B-array. Enhanced features such as wide-field imaging, automatic detection of source models, correcting for fast-switching source names, etc, are now also included for processing of any frequency over 1 GHz. This is the latest procedure known as VLARUN currently in AIPS (since March 2006). Further developments, also because of encountering different types of observations and user specified observing setups (and features..) are currently done in a private version until they are robust enough to be put back. This allows us to fiddle with the code, implement suggestions and, e.g., test out solar system mode without introducing bugs to VLARUN users that run the AIPS midnight job.

Current imperfections include non-robust automated self-cal and most added features have remained untested in spectral-line mode. No attempt is made to calibrate and image polarization data or mosaics.

VLARUN, the AIPS VLA pipeline procedure, was announced to the public at several occasions over the last few years, including the aforementioned "xraydio" meeting, the AIPS newsletters, and, e.g, the January 2005 AAS meeting in San Diego. Furthermore, VLARUN was used for several professional publications and explicitly mentioned in those articles.

2 The VLA pipeline setup

Although a description and manual of the fully automated VLA pipeline is the subject of another document (the "user manual" by Jared Crossley), a brief insight of the automated pipeline setup helps to understand the capabilities of the pipeline as described in the next section below.

The use of the VLA pipeline, VLARUN or our modified private version, can be divided up in two different modes. First, it can be run by a competent user that knows the data or has a drive to understand its details, and is capable of making decisions regarding, e.g., bad data, solution interval, image size, confusing sources, etc. Details about capabilities and input parameters are commented on below. Or, if a user is not familiar with the data, it can be run in a **best-effort** non-radio-expert mode, i.e., the pipeline can run without astronomical knowledge with the use of default "dummy" input parameters also outlined below. These two modes assume the user has started up AIPS and loaded the data into AIPS. Another mode, which is the automated pipeline mode, uses c-shell scripts, Perl programs and AIPS run files and default inputs analogous to the second "dummy" case.

All automated pipeline processing is done through command line instructions that interact with the NRAO archive, cron jobs and AIPS procedures. The operator of the automated pipeline only selects a data set to be processed and determines whether a run was successful. Other cron jobs and scripts automatically fill and index the image archive web pages (currently at http://www.aoc.nrao.edu/∼vlbacald). The operation is in three steps:

1: query the archive;

2: prepare a data set to be downloaded, checked, loaded into AIPS, and prepare a script to run the pipeline on it; and

3: after the pipeline has run validate the results and release the images and calibrated u, v-data to the NRAO and VLA image archive.

In order to run the automated pipeline smoothly, some preparations are done before the actual AIPS pipeline procedure is started. For example, all data sets are converted to J2000 coordinates, if needed, source and frequency checks are done and, if passed, to avoid confusion with another pipeline process each data set is processed under a separate AIPS user number. There are three different AIPS run procedure scripts:

a: prepare and check the set for pipelining (step 2 above);

b: if approved allow VLARUN to process it with standard parameters (without operator intervention after step 2 above); and

c: post-collect the images and data into a format standardized for release in the NRAO and VLA image archive (step 3 above).

3 The VLA pipeline in AIPS

This section describes the capabilities of the VLA pipeline procedure script VLARUN. Essentially it summarizes the VLARUN `HELP` file in AIPS....

VLARUN can be used to calibrate u, v-data and make investigative plots to refine the calibration, or it can be used to calibrate and automatically image the sources in one go. This switch is set by the user, specifying the parameter `DOIMAGES`. The request for images to be made by setting `DOIMAGES`>`0` activates an extra set of parameters for the imaging (see below), and setting `SLFCAL` to a non-zero value activates interactive self-calibration. The only other interactive setting is `NOPAUSE` which allows for a pause in the calibration process to check the flux densities and their associated errors of the secondary (phase) calibrators. It is very useful to look at these `''GETJY''` values to spot possible calibration problems (such as residual bad data). None of the other settings below modify the interactiveness of the pipeline procedure, although a request to make diagnostic plots (`AUTOPLOT`≥`0`) may slow down the pipeline considerably.

Note that many of the parameters are used in a non-intuitive way (i.e., not as usual in AIPS). Here the non-obvious parameters are described one-by-one, including their "dummy" value.

3.1 VLARUN calibration input parameters

`CATNUM` **or** `INNAME` (etc) specify the file to calibrate. For the automated pipeline this is always `WORKDISK=1`, `CATNUM=1`, and `INNAME=' '`, `INCLASS=' '` and `INSEQ=0`.

For spectral line data, `CATNUM` or `INCLASS` must point at the spectrally averaged `'CH 0'` file and a spectral line file `'LINE'` must exist with the same `INNAME` and `INSEQ`, and on the same `WORKDISK`.

`FASTSW=1` checks the source table for sources that are within 3 mas of each other, declares them the same source, and reassigns the shortest name found for them. This typically is needed when fast-switching is used at the VLA. The automated pipeline does this by default in an earlier stage (in the preparation stage 2 above), so for the automated pipeline `FASTSW=`−`1`.

If a flagging table is present it is assumed that removal of bad data has been done (either by hand in an earlier stage or in a previous run). Otherwise, setting `AUTOFLAG`≥`0` activates some automated flagging of bad data with `FLAGR`. If positive, `QUACK` is used to clip the beginning of each scan. However, for frequencies over 18 GHz this option is disabled to avoid clipping the majority of short scans - the user is warned to perform this outside of the pipeline if desired. For the automated pipeline `AUTOFLAG=2`, i.e., use all flagging options but disable `QUACK` for high frequencies. See also the section on improvements.

`PHAINT` is the parameter for the time interval used in `CALIB`, the core calibration task in AIPS. Typically `PHAINT` is short (a couple of minutes or a small fraction thereof for high-frequency observations) to obtain phase solutions. For amplitude solutions one would prefer to use a longer interval, which is implemented by specifying `AMPINT`. VLARUN first separately calibrates phases, then also the amplitudes. The automated pipeline uses `PHAINT=1` and `AMPINT=5` minutes.

One can specify a reference antenna with `REFANT`, but AIPS will make an educated choice if left zero as is done in the automated pipeline.

The `DOMODEL` switch allows `CALIB` to use AIPS provided source models to be used for calibration. This is usually good practice, though at this moment not all standard VLA flux calibration source models are available. Earlier versions of VLARUN would crash if a model was requested but did not exist; currently a point source (as was practice before the availability of these source models) is used in such cases. `DOMODEL` is set to use the models available in the automated pipeline.

If for some reason a standard flux calibrator is missing in the observation, the user can designate an alternate source observed in the observing run as point-like flux calibrator by setting `AMPCAL` to the source name, `FLUX` to its presumed flux density and when necessary `UVRANGE` to the range of *u,v*-values for which this flux density is defined. `AMPCAL` and `FLUX` are only used if no models are used (`DOMODEL`≤`0`), whereas `UVRANGE` may also apply to the default flux calibration models (though setting `UVRANGE` for a model is not recommended).

The `PHACAL` parameter captures the names of all phase calibrators to be used, up to 20. If more are needed or if the explicit calibrator names are unknown, as is the case in the automated pipeline, set `PHACAL='*',' '` and it will presume point sources for everything with a "calcode" in the source list. The latter usually works surprisingly well (for VLA data). Sources specified here are used as secondary amplitude (gain) calibrators and their flux densities will be determined by `GETJY`.

`BNDCAL` is only used in spectral line mode and captures up to five explicit names of bandpass calibrators to be used (simultaneously).

In interactive mode it is very useful to set `NOPAUSE=`−`1`. This will pause the pipeline just after `GETJY`, the task that derives the flux densities of the secondary phase calibrators. Next to the flux densities `GETJY` reports the errors in determining the flux densities, which together with the flux densities themselves may point out the success of, or residual bad data in, the calibration. In the automated pipeline this switch is turned off; `GETJY` flux densities can be examined in the automated pipeline log file.

The parameter `AUTOPLOT` is used to determine the amount of diagnostic plots. Setting this parameter however may slow the pipeline down considerably. Set `AUTOPLOT=`−`1` to skip diagnostic plots. The automated pipeline uses `AUTOPLOT=0` to make *u,v*-plots only for the calibrated single source files, i.e, the *u,v*-coverage, the visibility Real versus Imaginary and visibility amplitude versus *u,v*-distance. These *u,v*-plots are only made in the imaging stage (see below) when calibrated single source files are available. Positive settings will create plots during the calibration stage and include plots of phase and amplitude calibration solutions of the multi-source file and bandpass calibration solution in the case of a spectral line observation (see `HELP`).

Set `DOIMAGES`≤`0` to skip imaging, otherwise study the next two sections.

3.2 VLARUN imaging input parameters

These parameters are only of interest if imaging is requested (i.e., `DOIMAGES`>`0`).

The `ARRYSIZE` parameter fixes the resolution by taking the value as the maximum baseline length in kilometers. Alternatively, if zero it will figure out the maximum unprojected baseline from the antenna file, or, if set negative it will leave it up to the task `SETFC` to determine the image pixel size. The latter, using SETFC to determine the image pixel size, is done in the automated pipeline.

The first argument of `IMSIZE` specifies the output image size with a minimum of 128 (square). For calibrators this defaults to 256 and can only be disabled when imaging the full primary beams (as determined by `SETFC`). To do this set `IMSIZE` negative, or large negative to include full primary beam imaging for the calibrators in addition to the targets. The automated pipeline images the full primary beam for all sources.

NITER sets the maximum number of clean components to be used. Note that zero means zero, i.e., make a non-cleaned dirty image. Again, for calibrators this defaults to 500 (if ALLIMG>0 below), unless the full primary beam is imaged for the calibrators too (IMSIZE large negative). If set negative, as in the automated pipeline, NITER is set to a huge number effectively allowing CUTOFF (see next) to take over the stopping criterion.

If the image RMS noise reaches CUTOFF stop cleaning. This parameter heavily interacts with NITER. If CUTOFF is negative, stop cleaning after the first negative clean component (or NITER if it is reached first). If both NITER and CUTOFF are set negative, as is the case for the automated pipeline, the pipeline tries to make a conservative estimate for the RMS noise in the map and cleans to a conservative level above this estimated RMS noise level (and beyond the first negative), regardless whether this is a target source or a strong calibrator.

ALLIMG determines whether to treat the calibrators the same as the targets or not. Normally some information on the calibrators is desirable to obtain some quality assessment of the data. This can be an image (or total flux density, the source structure and position), though deep cleaning is not necessary. Set ALLIMG positive to include images of the continuum calibrators, larger than one for the spectral line version too and maybe also pay attention to NITER and CUTOFF. This mode, to make no distinction between calibrators and targets, is used in the automated pipeline.

3.3 VLARUN self-cal input parameters

This parameter is only relevant for radio continuum data of the target sources if imaging is requested (i.e., CATNUM **or** INCLASS point to an AIPS catalog entry with a class other than 'CH 0'/'LINE', and DOIMAGES>0).

The automated pipeline sets this parameter to zero, which at the moment is probably the best value. SLFCAL was included in an earlier version of the VLARUN code to attempt self-cal for strong point-like sources, allowing the task SCIMG to do the imaging instead of IMAGR. However, since the focus on the pipeline project and the many changes to the code, especially for continuum, it is unclear how everything has affected this option. No testing has been performed since the original implementation, so currently values other than zero are highly discouraged. The original implementation uses the absolute value of SLFCAL as the number of cycles (NMAPS in SCIMG) and the sign determines the level of interactive (TV) display (see HELP).

We hope to have an implementation for automated self-cal in the near future, though this may not use an adjusted or improved version of the current (SCIMG) implementation.

3.4 VLARUN pipeline parameters

If unknown or in doubt what to set for one or more of the parameters, try the automated pipeline "dummy" inputs. The results of using these inputs in the pilot automated pipeline imaging are described in a different document. The "dummy" inputs are below. Note that the ones with a † are different from the pipeline (as shown between brackets), but strongly recommended for interactive use. Also check your situation for the ones marked with ∗, as these are automatically imposed by the automated pipeline system:

BADDISK = 0	∗	AMPINT = 5		AUTOPLOT = 2 (0)	†
WORKDISK = 1	∗	REFANT = 0		DOIMAGES = 1	
CATNUM = 1	∗	DOMODEL = 1		ARRYSIZE = −1	
INNAME = ' '	∗	AMPCAL = ' '		IMSIZE = −1	
INCLASS = ' '	∗	FLUX = 0		NITER = −1	
INSEQ = 0	∗	UVRANGE = 0, 0		CUTOFF = −1	
FASTSW = −1	∗	PHACAL = '*', ' '		ALLIMG = 1	
AUTOFLAG = 2	∗	BNDCAL = ' '		SLFCAL = 0	
PHAINT = 1		NOPAUSE = −1 (1)	†		

4 Running the VLA pipeline

4.1 Starting conditions

In order for the pipeline to run smoothly it is important to pay attention not only to the input parameters, but also to the input u, v-file and its structure. For the automated pipeline, many of possible problems with the input u, v-file are dealt with in the preparation stage (2, or a), though it cannot handle every case. A simple scan listing using `LISTR` after loading the data into AIPS will indicate possible problems. Another problem saver is to use a flagging program such as `TVFLG` or `WIPER` to manually take care of bad data ahead of time as the automated flagging is not yet perfect. Also, currently all data at frequencies below 1 GHz (L-band) are disregarded by the pipeline.

Possible problems that can be spotted in the scan listing can be related to the source list or the observing setup itself. The pipeline cannot handle multiple frequency-ID's and calibration of multiple subarrays is untested. For each of these, split the data in single frequency files (`FILLM` does a lot of this already), and either reset all frequency-ID's to one (with `DFQID`) or split each frequency-ID (and subarray) in its own separate file. For spectral line data do this in the same way on both the `'CH 0'` and `'LINE'` data.

In the source list, look for a standard flux density calibrator (e.g., 3C286 which is also known as 1331+305, etc.) and whether it has enough visibilities; typically this should be a few integration times multiplied by the number of baselines. If data on all standard flux density calibrators are missing, find another source with enough visibilities for which the flux density information can be figured out - use this in `AMPCAL` and `FLUX`. Furthermore, verify whether all the phase calibrators have calibrator codes - this is usually the problem with phased-array data. The latter can be processed fine, though the setting `PHACAL='*',' '` cannot be used. Check that the calibrators have enough visibilities, and if possible, look up their approximate flux densities to compare with the `GETJY` values. Sometimes blocks of data may be missing due to weather or hardware problems. Be sure to be aware of any of such problems.

Bad data is best removed as soon as spotted to prevent the pipeline from running when bad results are to be expected. However, bad data can also be removed after calibration as it may be easier to recognize. If after the pipeline ran, it turns out that more data needs to be flagged, simply add the flags to the highest flagging table of the original input u, v-data set, delete the files that were created by the pipeline, and restart the pipeline. If a flagging table is present, automated flagging is disabled to retain the (manual) flagging. To enable the automated flagging either delete the flagging information (but this is probably not the intention), or apply the flagging by making a copy of the data with the flagged data removed. Note also that for high frequencies (i.e., over 18 GHz), the automatic `QUACK` flagging is disabled. If needed, run `QUACK` manually and if other automated flagging is requested, make a copy of the file with flagging applied.

4.2 Suggested use

The automated pipeline runs VLARUN blindly with the parameters set as described in Section 3. Most of the time there are no problems, but some exceptions are mentioned below. However, since the pipeline runs very fast it is not a big waste of time to run it again, especially if one is already working within AIPS. Because the pipeline relies on a sequence of commands, the AIPS prompt is inactive during the pipelining process - it is useful to open up another window with AIPS running using the same AIPS user number (see below).

A personal recommendation is to do the following; this is just an empirically tested mode of operation aimed to obtain a high success rate on a random data set - specific data sets may run with less, or perhaps more complicated schemes. After loading and quickly inspecting the data scans (but no flagging yet), run the pipeline with automatic flagging enabled (`AUTOFLAG`$\geq$`1`), with `NOPAUSE=`$-$`1`, and without imaging requested (`DOIMAGES=`$-$`1`). Check which antenna is used as reference antenna and check whether this is proper. This will yield first order flagging, and fluxes with errors for the phase calibrators. Match large errors with additional flags in the highest numbered flagging table, and delete the files created by the pipeline (which should be a single TASAV-class file only). It is suggested to leave the pipeline window for the pipeline and to use the other window for these non-pipeline operations.

Rerun the pipeline with some plotting options (e.g., AUTOPLOT=2), perhaps with a different reference antenna and perhaps with a modified solution interval. Request imaging of a reasonable field of view, with a reasonable (low) number of iterations and a reasonable estimated cutoff flux set. Inspect the calibration plots and images generated to determine phase stability, residual bad data, (calibrator) source structure, etc. Delete the files created by the pipeline (a single 'TASAV'-class file, and several u, v-files, e.g., with class 'X-BAND', and corresponding image files with class 'ICL001').

Finally rerun the pipeline in the way the final images should be. One can always clean deeper afterward by rerunning IMAGR if NITER and CUTOFF were set too conservatively. If SETFC is used, make sure to make a copy of the BOXFILE before IMAGR finishes. At the time of imaging the target, a simple "tget imagr" in the additional AIPS window may be very useful for later execution with an adjusted number of iterations, field size, etc..

5 Common problems and fixes

Most problems relate to either bad data, bad observing conditions, or bad pipeline constraints. In particular the automated flagging is not yet perfected and occasionally obviously bad data may not be flagged automatically. However, bad data is easily removed by interactive flagging programs and rerunning the pipeline with additional flagging information in the flagging table is trivial.

In some cases the automated flagging deletes too much data, in particular in fast-switching observations. Or, when slewing of the telescope leaves too little data on the first (calibrator) scan, flagging can remove most or all of the data of those scans. If not enough data remains, calibration solutions may be too noisy or undetermined. Remove the flagging table, flag obvious bad data by hand, maybe apply the flags (Section 4), and rerun the pipeline.

Other problems may occur because the input u, v-data file does not have the expected structure. For example, none of the standard flux calibrators observed has enough visibilities, the data contains multiple frequency-ID's or subarrays, or calibrator codes are missing. Make sure the starting conditions are "clean" (see above). Occasionally unused qualifiers are attached to the source, which can be reset using DQUAL. Nothing can be done to correct for "unfortunate" observing conditions such as weather, bad pointing, faulty hardware and pilot errors.

Finally, sometimes the pipeline gets confused when files exist with similar names (but, e.g., only with a different sequence). Make sure before starting the pipeline that no files remain from a previous run - rename or delete them to avoid confusion.

Be aware that in a few cases, a data set just cannot be calibrated, period! If a fix still can't be found, try to email lsjouwerman@aoc.nrao.edu for help.

6 Improvements

Currently, improvements are implemented in the next version of VLARUN in the area of self-cal algorithms and dynamic range problems, as well as processing of "moving target" (i.e., solar system) data. Furthermore, it should be possible to use more than one standard flux calibrator, or a model of the source specified in AMPCAL. Longer time-scale improvements are expected to include more robust flagging and perhaps polarization calibration. It is also possible to check and update calibrator positions and calibrator codes on the fly and perhaps create models for the calibrators prior to the target calibration. However the priority and implementation of part or all of this depends on local priorities.

Acknowledgments

Many thanks go to Eric for making some of my wishes come true, and to Amy, John, Ed and Jared for all their suggestions and other input.