

Research Paper

Calibration database for the Murchison Widefield Array All-Sky Virtual Observatory

Marcin Sokolowski¹, Christopher H. Jordan^{2,3}, Gregory Slep², Andrew Williams², Randall Bruce Wayth^{1,3}, Mia Walker², David Pallot⁴, Andre Offringa^{5,6}, Natasha Hurley-Walker¹, Thomas M. O. Franzen⁵, Melanie Johnston-Hollitt¹, David L. Kaplan⁷, David Kenney¹ and Steven J. Tingay¹

¹International Centre for Radio Astronomy Research, Curtin University, Bentley, WA 6102, Australia, ²Curtin Institute of Radio Astronomy, GPO Box U1987, Perth, WA 6845, Australia, ³ARC Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), Australia, ⁴International Centre for Radio Astronomy Research, University of Western Australia, Crawley 6009, Australia, ⁵Netherlands Institute for Radio Astronomy (ASTRON), 7991 PD, Dwingeloo, The Netherlands, ⁶Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV, Groningen, The Netherlands and ⁷Department of Physics, University of Wisconsin–Milwaukee, Milwaukee, WI 53201, USA

Abstract

We present a calibration component for the Murchison Widefield Array All-Sky Virtual Observatory (MWA ASVO) utilising a newly developed PostgreSQL database of calibration solutions. Since its inauguration in 2013, the MWA has recorded over 34 petabytes of data archived at the Pawsey Supercomputing Centre. According to the MWA Data Access policy, data become publicly available 18 months after collection. Therefore, most of the archival data are now available to the public. Access to public data was provided in 2017 via the MWA ASVO interface, which allowed researchers worldwide to download MWA uncalibrated data in standard radio astronomy data formats (CASA measurement sets or UV FITS files). The addition of the MWA ASVO calibration feature opens a new, powerful avenue for researchers without a detailed knowledge of the MWA telescope and data processing to download calibrated visibility data and create images using standard radio astronomy software packages. In order to populate the database with calibration solutions from the last 6 yr we developed fully automated pipelines. A near-real-time pipeline has been used to process new calibration observations as soon as they are collected and upload calibration solutions to the database, which enables monitoring of the interferometric performance of the telescope. Based on this database, we present an analysis of the stability of the MWA calibration solutions over long time intervals.

Keywords: astronomical databases: miscellaneous – instrumentation: interferometers – methods: data analysis – techniques: interferometric – virtual observatory tools

(Received 19 April 2020; revised 28 April 2020; accepted 4 May 2020)

1. Introduction

The Murchison Widefield Array (MWA; Tingay et al. 2013; Wayth et al. 2018) is one of the low-frequency precursors of the Square Kilometer Array (SKA^a). It is located at the Murchison Radio-astronomy Observatory (MRO), the future site of the low-frequency component of the SKA (SKA-Low). The MWA began its operations in 2013 and since then has recorded over 34 petabytes of visibility (output from the MWA correlator) and voltage capture system (Tremblay et al. 2015) data, which are archived in the Pawsey Supercomputing Centre (PSC). Based on the MWA Data Access Policy^b, the data become publicly available 18 months after collection or immediately after collection for the members of

the MWA collaboration. As a consequence, most of the archival visibility data (approximately 19.3 PB, representing 79% of the total) are now (as of March 2020) publicly available. Public data have been available via the MWA All-Sky Virtual Observatory (MWA ASVO^c) since its initial pilot release in 2017. The pilot MWA ASVO interface enabled users to download raw MWA data in standard radio astronomy data formats such as CASA measurement sets (McMullin et al. 2007) or UV FITS files (Greisen 2019). The data sets returned to the end user are flagged for radio-frequency interference (RFI) using AOFLAGGER software (Offringa et al. 2012; Offringa et al. 2015) and averaged in time and frequency according to the requirements specified in the web interface or request file. However, these data are not calibrated and require several additional steps before sky images could be formed, which could have been difficult for users not familiar with MWA data calibration. Here, we present a calibration extension of the MWA ASVO, opening a new avenue for any researcher worldwide without deep knowledge of the details of the MWA instrument and data processing to download calibrated visibility data in the aforementioned formats.

Author for correspondence: Marcin Sokolowski, E-mail: marcin.sokolowski@curtin.edu.au

Cite this article: Sokolowski M, Jordan CH, Slep G, Williams A, Wayth RB, Walker M, Pallot D, Offringa A, Hurley-Walker N, Franzen TMO, Johnston-Hollitt M, Kaplan DL, Kenney D and Tingay SJ. (2020) Calibration database for the Murchison Widefield Array All-Sky Virtual Observatory. *Publications of the Astronomical Society of Australia* 37, e021, 1–10. <https://doi.org/10.1017/pasa.2020.17>

^a<https://www.skatelescope.org/>

^b<http://www.mwatelescope.org/team/policies>

^c<https://asvo.mwatelescope.org>

The calibration database (CALDB) has been populated with calibration solutions from the entire history of MWA observations. In order to do it, we developed a dedicated calibration pipeline. The newly collected calibration observations are automatically processed in near-real time, and the resulting calibration solutions are uploaded to the CALDB. This enables us to monitor the stability of their phase and amplitude components, that is, the interferometric performance of the MWA, which allows the MWA Operations team to identify problems which may go undetected by other components of the Monitor and Control (M&C) software^d.

This paper is organised as follows. In Section 2, we present the structure of the CALDB. In Section 3, we describe software pipelines developed to populate the database with calibration solutions derived from archived calibrator observations since 2013, and a near-real-time version of this pipeline which is used to process new calibrator observations (collected just after sunset and before sunrise). We also present a system developed for handling requests for missing calibration solutions, and web services developed to download calibration solutions for a specific observation via web browser or command line (WGET command). In Section 4, we describe the applications of the CALDB, such as the MWA ASVO, monitoring of the interferometric performance of the MWA, providing calibration solutions to the new MWA correlator, and potential future applications for transient and ionosphere monitoring. Finally, in Section 5, we make summarising remarks and discuss the importance of these developments in the context of the future SKA-Low telescope.

2. Database of calibration solutions

The overview of the current MWA ASVO system is shown in Figure 1. The newly added calibration component consists of the CALDB, and several scripts and pipelines for: populating this database with calibration solutions, accessing the database, and applying solutions to uncalibrated data downloaded from the MWA archive at the PSC.

2.1. Database structure

The CALDB has been implemented as a PostgreSQL database^e, which is an advanced open-source relational database and has also been used for storing other M&C MWA data. Presently, the MWA can record up to 30.72 MHz of bandwidth split into 24 coarse channels of 1.28 MHz each. Figure 2 shows phase of calibration solutions in the frequency range 70–230 MHz computed from Pictor A observations recorded between 10:32 and 10:48 UTC on the 2020 March 24. This figure shows that phase as a function of frequency is very well modelled by a linear function as unaccounted delays (due to cables or fibres) are the main contributors to the MWA calibration terms. Therefore, fitted parameters of a linear function provide a compact and efficient way of storing phase of calibration solutions, which is also robust against any ripples caused by reflections in cables (e.g., Tile 051 in Figure 2 or Figure 4) or inaccuracy of the sky model used in the calibration process. This way four double precision values (two for each polarisation) are preserved for each tile^f instead of two times the

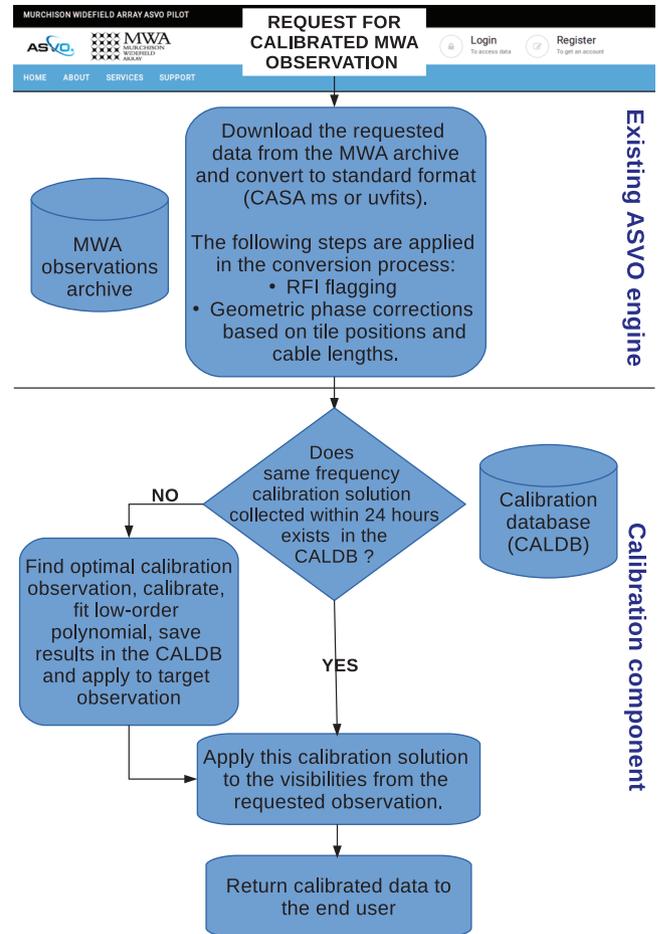


Figure 1. General overview of the MWA ASVO system with the new calibration component.

number of fine channels (typically 768). Amplitude, on the other hand, can have complicated frequency structure related to the MWA tile bandpasses. However, locally (in a sufficiently narrow frequency band), it can also be approximated by a linear function; and the natural choice of the narrow band for the linear fit is the 1.28 MHz MWA coarse channel. Therefore, the database was designed to store parameters of low-order (first-order) polynomials fitted to amplitudes and phases of calibration solutions as a function of frequency. The original calibration solutions are stored on a hard drive and are not inserted into the database in order to keep the database compact. We envisage that this approach will likely continue in the future. The CALDB is a part of the MWA M&C schema and consists of the following three tables (Figure 3):

- **calibration_fits**: it provides versioning of calibration solutions stored in the database. It enables uploading newer (possibly better) calibration solutions without the necessity of removing the older versions from the database. The table contains its own unique identifier (FITID field), a reference to observation identifier (OBSID field), and a timestamp of calibration solution (FIT_TIME field).
- **calibration_solutions**: each record of this table contains information about calibration solutions for both polarisations (X and Y) of a single MWA antenna. Besides references to CALIBRATION_FITS record (FITID field) and the observation identification field (OBSID), it contains calibration fields for both

^d<http://www.mwatelescope.org/telescope/monitor-control>

^e<https://www.postgresql.org/>

^fMWA 16-dipoles antenna units are commonly referred to as 'tiles'.

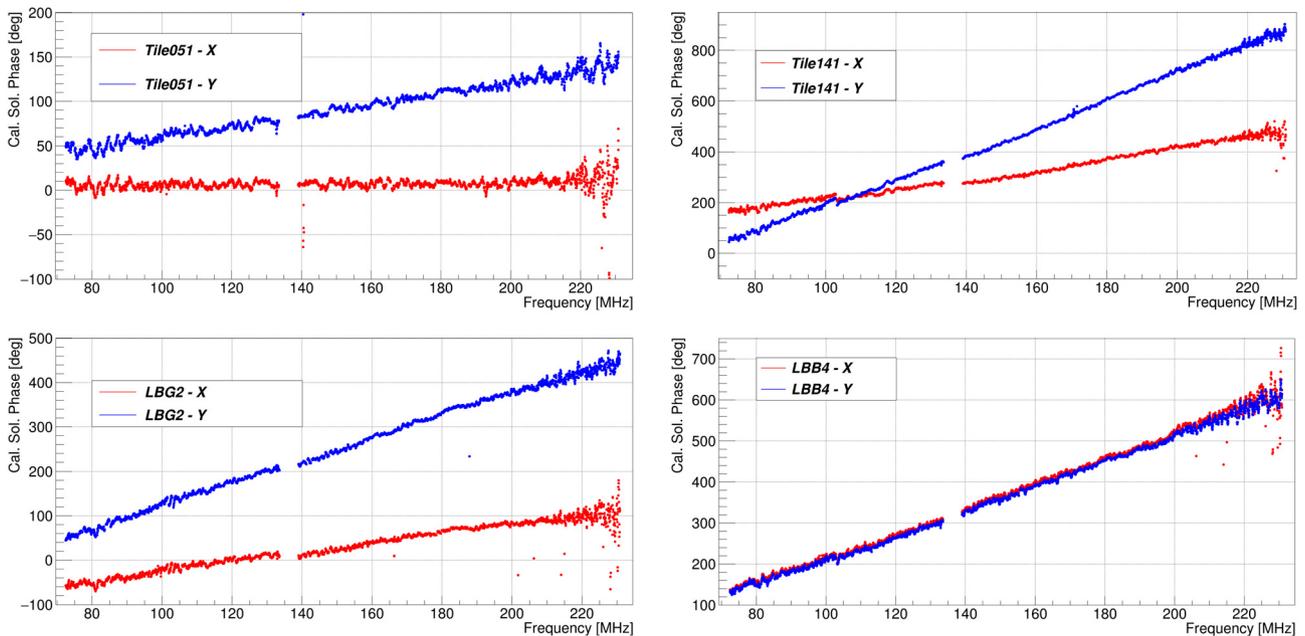


Figure 2. Calibration solution phase as a function of frequency in the range 70–230 MHz for selected MWA tiles (the MWA 16-dipole units are commonly referred to as ‘tiles’) in X and Y polarisations. The data were recorded on 2020 March 24 between 10:32 and 10:48 UTC. Tiles LBB2 and LBG2 are long-baseline MWA tiles. These figures show that the phase is very well modelled by a linear function of frequency over the full MWA band. The slope of a given tile depends on the electrical length of its signal path, which can be slightly different in X and Y polarisations (Tiles 051, 141, and LBG2). Frequencies above 220 MHz are affected by radio-frequency interference.

polarisations with names differing by prefix in field names (x_{-} or y_{-}). Slope (fields $x_{\text{delay_m}}$ and $y_{\text{delay_m}}$) and intercept (fields $x_{\text{intercept}}$ and $y_{\text{intercept}}$) fitted to the phase of calibration solutions are used to describe the phase of the calibration solution over the entire 30.72 MHz of the MWA’s instantaneous bandwidth, which can be either continuous or non-continuous (the latter is commonly called ‘picket fence’ mode). The fitted slopes are converted to length units using speed of light in vacuum and these values are stored in the database. First-order polynomial fits (thus two database fields) are sufficient to accurately fit the phase over the full observing band provided an accurate sky model is used in the calibration process (this will be described in Section 3.2), which was verified during development and testing. Amplitudes of calibration solutions were also fitted with a linear function, but in this case the fit was performed over every 1.28 MHz coarse channel. Therefore, they are stored as two arrays (for X and Y polarisations) of real values (typically 24, but the arrays are of variable size). This table also contains quality flags, which are real values in [0, 1] range ($X_{\text{PHASE_FIT_QUALITY}}$ and $X_{\text{GAINS_FIT_QUALITY}}$ for X polarisation). These flags are calculated as ratios of the number of ‘good’ channels, where the difference between the original value of a calibration solution (either phase or amplitude) and the fitted curve is smaller than five standard deviations (5σ), to the number of all channels. Ratio values above 0.6 are considered good quality calibration solutions.

- **cal_solution_request:** it enables requests for missing calibration solutions. If the user requests calibrated data which do not have corresponding calibration solutions for the same frequency channels within 12 h in the CALDB, a new request record is inserted into the table `CALSOLUTION_REQUEST` (if it is not already present there). Then an automatic script finds all the new records in this table, identifies corresponding calibration

observations in the main MWA database, calibrates them, and uploads resulting calibration solutions into the CALDB. If the appropriate calibration observations cannot be found or the calibration procedure fails, the error message is stored in the field `ERROR` and can be returned to the end user.

2.2. Present status of the database

Currently (as of March 2020), the database contains calibration solutions from around 11 200 calibration observations, which provides, on average, five calibration solutions per day; one for each of the primary MWA frequency bands (at centre frequencies of 88.32, 119.04, 154.88, 185.6, and 216.32 MHz). The database grows every day as new calibrator observations are collected and the near-real-time pipeline calibrates them and inserts calibration solutions into the database.

In order to populate the database with the historic and new calibration solutions, we developed a dedicated reduction pipeline, described in Section 3.

3. Automatic calibration pipelines

Originally, the pipeline used CASA software to calibrate calibrator observations in near-real time and create control images of the calibrator observations. In order to calibrate many archival calibrator observations, we developed a new pipeline using software more suited to the MWA observations. We are upgrading the current contents of the database with calibration solutions from the new pipeline in order to create a uniform database of calibration solutions resulting from the same data reduction pipeline, software, and sky model. Both pipelines use the MWA ASVO interface to download uncalibrated CASA measurement sets of calibrator observations, which are produced on the MWA ASVO servers. As

Database column name	CALIBRATION_FITS <i>describes an instance of calibration solution fit for a given observation ID (obsid field)</i>
fitid	Primary key (internal ID)
obsid	MWA observation ID
code_version	Version of code used
fit_time	Timestamp of the fit
creator	Creator of the fit
fit_niter	Number of fitting iterations
fit_limit	Fit convergence limit

Database column name	CALSOLUTION_REQUEST <i>Records automatically added when ASVO user requests observation with missing calibration solution</i>
obsid	MWA observation ID
unixtime	Timestamp of record creation
status	Status of request processing : 0 - new request to be processed 1 - being processed 2 - completed without errors -1 - error occurred
error	Error description
obsid_target	Target observation ID to which calibration will be applied

Database column name	CALIBRATION_SOLUTIONS <i>Table containing fitted calibration solutions for all MWA antennas (tiles) for a given observation ID (obsid field)</i>
fitid	Primary key (internal ID)
obsid	MWA observation ID
tile_id	MWA tile (antenna) ID (defined in metadata tables)
x_delay_m	Fitted delay in metres for X polarisation
y_delay_m	Fitted delay in metres for Y polarisation
x_intercept	Fitted intercept in metres for X polarisation
x_gains []	Arrays of mean gains for X and Y polarisations (per MWA coarse channel)
y_gains []	Arrays of mean gains for X and Y polarisations (per MWA coarse channel)
x_gains_pol0 []	Arrays of the fitted 1 st order polynomial parameters for X pol. for each coarse channel
x_gains_pol1 []	Arrays of the fitted 1 st order polynomial parameters for X pol. for each coarse channel
y_gains_pol0 []	Arrays of the fitted 1 st order polynomial parameters for Y pol. for each coarse channel
y_gains_pol1 []	Arrays of the fitted 1 st order polynomial parameters for Y pol. for each coarse channel
x_phase_sigma_resid	Standard deviations of residuals of the 1 st order polynomial fit of phase across the observing band (quality indicator)
y_phase_sigma_resid	Standard deviations of residuals of the 1 st order polynomial fit of phase across the observing band (quality indicator)
x_phase_chi2dof	Chi2 / NOF of the 1 st order polynomial fit of phase across the observing band (quality indicator)
y_phase_chi2dof	Chi2 / NOF of the 1 st order polynomial fit of phase across the observing band (quality indicator)
x_phase_fit_quality	Value from range 0 -1 representing quality of the phase fit in X and Y polarisation. It is a ratio of number of fine channels with distance to the fit smaller than 5 standard deviations to a number of all channels. Typically values above 0.75 mean a good quality fit.
y_phase_fit_quality	Value from range 0 -1 representing quality of the phase fit in X and Y polarisation. It is a ratio of number of fine channels with distance to the fit smaller than 5 standard deviations to a number of all channels. Typically values above 0.75 mean a good quality fit.
x_gains_sigma_resid	Arrays of standard deviations of residuals of the 1 st order polynomial fit of gain amplitude across a single coarse channel (quality indicator)
y_gains_sigma_resid	Arrays of standard deviations of residuals of the 1 st order polynomial fit of gain amplitude across a single coarse channel (quality indicator)
x_gains_fit_quality	Arrays of fit quality flags (values in 0 - 1 range) for X and Y polarisations - for explanation see description of x_phase_fit_quality field (above)
y_gains_fit_quality	Arrays of fit quality flags (values in 0 - 1 range) for X and Y polarisations - for explanation see description of x_phase_fit_quality field (above)

Figure 3. The tables in the MWA ASVO calibration database.

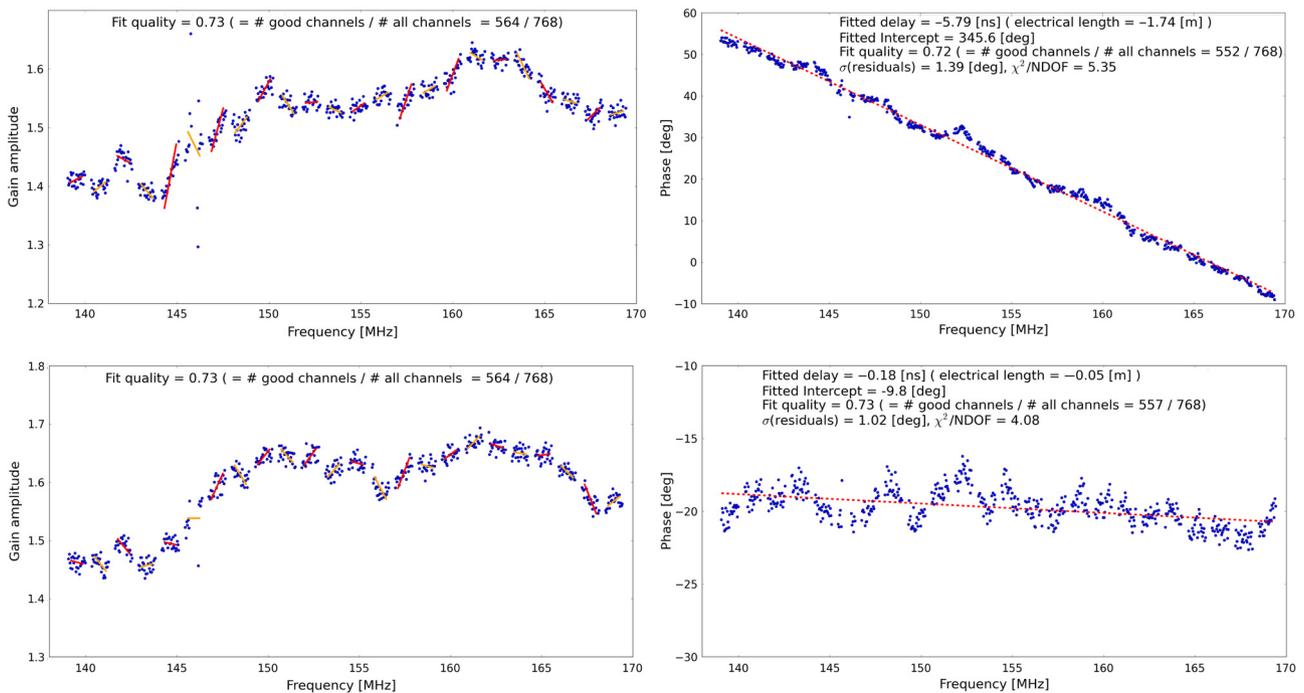


Figure 4. Examples of calibration solutions fitted with first order polynomials for MWA tiles 136 (upper row) and 065 (lower row). These type of plots are automatically generated whenever new set of calibrator observations are collected (just before sunrise and shortly after sunset). Left panels show examples of first order polynomials fitted to the gain amplitude in each of the 24 MWA coarse channels. Right panels show examples of linear functions fitted to phase over the MWA continuous observing band of 30.72 MHz. It can be noted that amplitude fitting may not work well for coarse channels affected by radio-frequency interference (RFI), and a combination of wider band fitting with more fit/exclude outliers iterations can further improve the final quality of the fitted amplitudes of calibration solutions.

described in Section 3.5, the COTTER program is used in the conversion process, which implies that RFI flagging is also applied at this stage.

3.1. Near-real-time calibration of new calibration observations

The CASA-based pipeline has been used to reduce newly collected MWA calibrator observations. Every day, the MWA observes a calibrator source shortly after sunset and before sunrise at five standard frequency bands (at centre frequencies of 88.32, 119.04, 154.88, 185.6, and 216.32 MHz) and in the so-called ‘picket fence’ mode with 24 coarse channels spread regularly over the frequency range 78–240 MHz. The calibrator script continuously runs on one of the MWA servers, checks the MWA schedule database for new calibrator observations and whenever it detects that new calibrator observations were collected they are automatically downloaded, calibrated and control images of the calibrator field are formed at selected frequencies. Presently, the observations are downloaded from the PSC and processed on a server at Curtin Institute of Radio Astronomy (CIRA), which introduces additional delay. In the future, this processing will be relocated to the MRO, and the pipeline will use ‘raw’ visibility files as they are produced by the MWA correlator. If the quality of the resulting calibration solutions satisfies minimum requirements, they are uploaded to the CALDB. The requirements for the new calibration solutions to be loaded to the database are the following: (i) they must be better than the ones already in the database (if there are any for this OBSID) and (ii) more than half of antennas have acceptable calibration solutions, where ‘acceptable’ means that more than 60% of the channels have a phase fit within 5σ from the data (this criteria is to avoid storing calibration solutions of very low quality). These near-real-time calibration solutions are used for monitoring of the interferometric performance and enable us to examine the long-term stability of the MWA (Section 4).

3.2. Calibration solutions of archived data

The CALDB has been populated with calibration solutions starting from the beginning of 2013. In order to achieve this, we created a list of all calibrator observations and submitted them for processing by the calibration reduction pipeline HERACLES[§]. The pipeline is using CALIBRATE software (Offringa et al. 2016) upgraded with the newest 2016 MWA beam model (Sokolowski et al. 2017) and sky model generated by PUMA^h (Line 2018). It creates binary files with calibration solutions and control images of the field using the WSCLEANⁱ program (Offringa et al. 2014).

3.2.1. HERACLES pipeline

Our initial attempt for calibration of the MWA archive used a single virtual machine on the cloud-based system ‘Nimbus’ hosted by the PSC. However, our task quickly proved to be insufficient for the amount computing resources afforded by this system. For this reason, HERACLES was converted into a generalised and distributed system, which could be run on any free resources available (such as unused desktop computers in CIRA).

HERACLES is primarily utilised by a single executable which has two modes of operation: server and client. The server mode is primarily concerned with which observations need to be calibrated, based on a SQLITE database. This database tracks which observations have not yet been calibrated, which have been calibrated, and which have failed. The HERACLES server must also coordinate with the state of any MWA ASVO data downloads (available, in progress, not available, etc.). So as to not flood the MWA ASVO with download requests, the HERACLES server uses a runtime setting to prepare a certain number of observations for download as the clients progress.

Once connected to a server, the operation of a HERACLES client follows a simple loop:

- (i) request an observation to calibrate. If an observation is ready, move to step (ii), otherwise, wait for 1 min before querying the server again;
- (ii) download the observation;
- (iii) operate upon the data with an executable (set at runtime, typically a BASH script);
- (iv) if the result of the executable was a success, the calibration solutions and any other useful products are transmitted to the server. Return to step (i); and
- (v) if any failure occurs in the loop, it is also reported to the server, before returning to step (i).

As the computational load of the server is negligible, clients may also be run on the same computer as the server.

The mode of operation of HERACLES clients allows users to enable or disable clients dynamically, which allows otherwise unused computing resources to be utilised, and proved to be an effective, efficient solution for calibrating a large volume of data. Within a few months, we were able to download, calibrate, and image observations and insert solutions into the CALDB from nearly 6 yr of MWA operations.

3.2.2. CASA pipeline

Originally, the CASA-based pipeline was used to calibrate evening and morning calibration scans and store the calibration solutions in the database. This pipeline used VLA Low-Frequency Sky Survey Redux [VLSSr; Lane et al. (2014)] images of calibrator sources (such as Hydra A, 3C444, Hercules A, and Pictor A) to derive calibration solutions. Since the creation of the new HERACLES pipeline, the CASA pipeline will be retired and the calibration solutions in the database are being superseded with the results from the new pipeline.

3.3. Uploading calibration solutions to the database

Phase and amplitude of calibration solutions resulting from the reduction pipeline are fitted with the first-order polynomial as a function of frequency. Figure 2 shows that phase of calibration solutions is a linear function of frequency over a very wide band. The lowest and highest four 40 kHz fine channels (160 kHz) in each coarse channel as well as the fine channels flagged due to RFI (during the conversion process) are excluded from the fitting.

First, the phase of calibration solutions over the 30.72 MHz band is ‘unwrapped’; phase values are not limited to $[-180, +180]$ degrees, but can range from minus infinity to plus infinity. Then the phase is fitted with a linear function over the

[§]<https://gitlab.com/chjordan/heracles>

^h<https://github.com/JLBlLine/srclists>

ⁱ<https://sourceforge.net/p/wsclean/wiki/Home/>

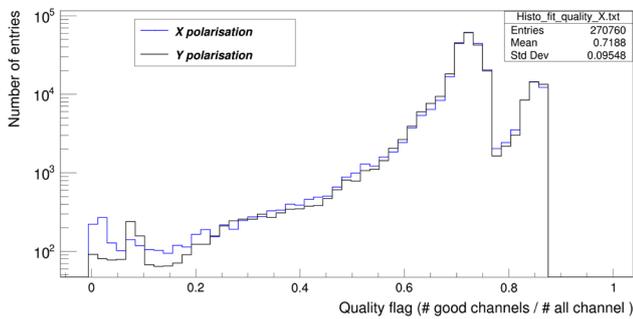


Figure 5. Distribution of quality flag for calibration solutions in the MWA ASVO database for X and Y polarisations. Quality flag does not exceed ≈ 0.88 because fine channels at the edges of coarse channels are always flagged. The dual peak structure is due to the fact that calibration solutions in coarse channel 121 (154.88 MHz), which is the optimal MWA operating frequency, are noticeably better than at other frequencies and they generate a peak at around 0.85. There are 419 observations with flag quality exceeding 0.75 for all antennas (in both polarisations) and these are observations of: 3C444 (224), Pictor A (59), Centaurus A (45), Hydra A (29), PKS2356-61 (27), and a drift scan field (11), where the numbers in brackets are the number of observations for the particular calibrator. This indicates that majority of best quality calibration solutions comes from observations of 3C444. The main peak at around 0.73 is due to good quality calibration solutions at other standard frequencies (88.32, 119.04, 185.6, and 216.32 MHz).

entire observing band (30.72 MHz in continuous observations) resulting in two fit parameters: slope and intercept (right column in Figure 4). These are sufficient to accurately describe the phase of calibration solutions as a function of frequency, provided that the sky model used in the calibration process is complete (this was verified in the development and testing stage). The fitted slope is converted to a corresponding time delay (Δt) and eventually the length $c\Delta t$ (in metres), where c is speed of light in vacuum, is saved to the CALDB.

The amplitude of calibration solutions is also fitted with the first-order polynomial and in this case the fit is performed over a single 1.28 MHz coarse channel, resulting in different slopes and intercepts for each of the 24 coarse channels (left column in Figure 4). It was verified that linear fit is the optimal polynomial order to fit amplitudes over an MWA coarse channel as a parabola had only slightly lower χ^2 values and nearly two times higher Bayesian information criterion value (Schwarz 1978), which proved that the linear fit is a more appropriate representation of data than the parabola.

If the fit satisfies quality requirements, the resulting fitted parameters are stored in the database. The current quality requirement is that the ratio between number of good quality channels to all the channels in the calibrated observation is above 0.6 (Figure 5).

3.4. Accessing calibration solutions in the database

The calibration solutions in the database can be accessed via a web service with a standard `WGET` command^l. The request is executed on an MWA server, and if the appropriate (the same frequency band and within 24 h from the target observation) calibration solution exists in the database, it is returned to the user in the

same binary file format (‘.bin’ file) as produced by the calibration procedure developed by Offringa *et al.* (2016). If there is no suitable calibration solution, an error message stored in a text file is returned to the end user.

3.5. Application of calibration solutions to data downloaded from the MWA ASVO

The MWA ASVO website and API^k allow users to submit ‘conversion’ jobs which, when run, retrieve the observation, pre-process the data, converting the raw MWA correlator visibility format into a standard CASA or UV fits format, and then make the data product available for download. The conversion/pre-processing steps expose the options available in the `COTTER` pre-processing pipeline (Offringa *et al.* 2015). The MWA ASVO calibration option utilises a recently added feature of `COTTER`, allowing calibration solutions retrieved from the CALDB to be applied to the data before any data averaging takes place.

In a typical conversion job with the calibration option set, the requested observation is staged from the Pawsey Long-Term Archive (LTA). The LTA has a hierarchical storage management system consisting of several different tiers of storage, ranging from 1.5 PB of spinning disc to an allocation of 40 PB of magnetic tape. Once the observation data are available on the disc cache, they are copied to a scratch area. A web service call is made to retrieve the metafits file that contains much of the metadata associated with the requested observation—this is also stored with the observation files in the scratch area.

The calibration web service is called by the MWA ASVO to retrieve the best calibration solution for this observation (Section 3.4). If the solution is found, then the calibration solution binary file is retrieved and stored with the rest of the observation files in the scratch area. If no calibration solution is suitable, then the job fails and a request record is added to the CALDB to produce a solution for this observation. The user is informed to try again once this is complete (usually within 24–48 h).

With all of the files now available, the server then executes `COTTER`, with the ‘-full-apply’ command line argument, which applies the provided calibration solution to every integration before averaging (if requested). There is also another new `COTTER` option ‘-apply’ which applies provided calibration solutions after averaging integrations over a requested interval. Once `COTTER` has produced the output data in a standard radio astronomy data format, a download URL is provided to the user via the website or API so the data can be retrieved. During the conversion process, RFI flags (either pre-computed by `AOFLAGGER` or calculated by `COTTER`) are also applied to the data. Hence, the resulting data sets do not require any further pre-processing, and initial sky images can be formed using standard radio astronomy software tools, such as for example `WSCLEAN`, `CASA`, or `MIRIAD` (Sault, Teuben, & Wright 1995). These initial images can be used in self-calibration procedure in order to improve calibration solutions and/or further processing steps, such as primary beam^l or ionospheric corrections^m, may be applied depending on the requirements of the specific science case.

^l`WGET` http://mro.mwa128t.org/calib/get_calfile_for_obsid?obs_id=OBSID&zipfile=1&add_request=1 where OBSID should be replaced with the observation ID of the target or explicit calibrator observation.

^k<https://github.com/ICRAR/manta-ray-client>

^lhttps://github.com/MWATelescope/mwa_pb

^mhttps://github.com/nhurleywalker/fits_warpy

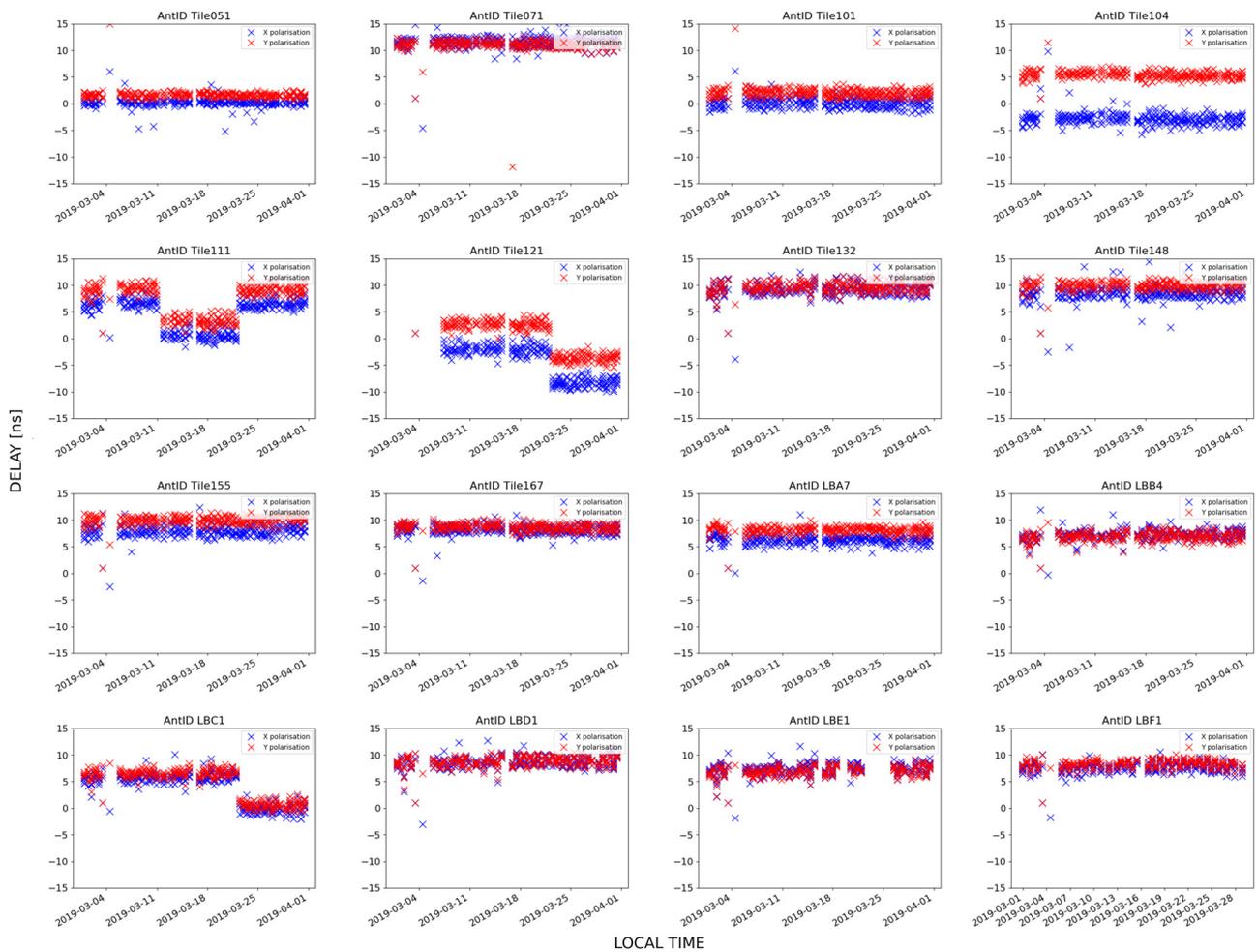


Figure 6. Delay (ns) fitted to the phase of calibration solutions as a function of time, in March 2019 when the MWA was in the extended configuration (Wayth et al. 2018). Sixteen tiles representing typically observed behaviour are shown. The sudden ‘jumps’ by approximately 6 ns (for Tiles 111, 121, and LBC1) were due to receiver resets (affects all eight tiles connected to a particular receiver). The delay usually returns to the pre-reset value after the next receiver reboot. Because of this effect, the cable delays are usually not adjusted to an accuracy better than 4 m (1 m corresponds to time delay of ≈ 3.33 ns). It can be seen that when receivers are not rebooted, the telescope remains very stable over timescales of weeks.

4. Other applications of the calibration pipelines and database

Besides the main application of the CALDB, which is to enable downloading of calibrated data in standard astronomy data formats via the MWA ASVO interface, there are several other benefits of having a complete database of calibration solutions, which will be described in this section.

4.1. Monitoring performance and stability of the MWA telescope

The near-real-time pipeline reduces daily calibrator observations, fits their phases and amplitudes with first-order polynomials (Figure 4), and inserts the resulting cable delays and intercepts into the CALDB. These fitted cable delays can be plotted as a function of time to monitor the long-term stability of the instrument. If the system is stable, the slope should be approximately constant over long periods of time (timescales of weeks or even months). Figures 6 and 8 show fitted delays (in nanoseconds) as

a function of time for selected 16 MWA tiles in the extended and compact configurations, respectively. It can be seen that the instrument remains very stable over many weeks. A compilation of such plots for all tiles is shown in Figures 7 and 9 where standard deviation of fitted delays is plotted against the antenna index enabling aggregation of the system stability in a single plot.

Routine monitoring of these plots enabled the identification of problems which can remain undetected in real-time plots showing power spectra of all the MWA tiles. In particular, it enables the monitoring of clock signals connected to the MWA receivers and in a few cases it identified the ‘drift’ of a receiver clock due to a failure at the initialisation process, which was fixed by rebooting the receiver. It was also noticed that the phases of the calibration solutions can abruptly change when an MWA receiver is power-cycled and the clock latches with an accuracy of 10 ns resulting in a step-like change of slope (corresponding to less than 3 m of length using speed of light in vacuum). Since this is not a large delay, with insignificant impact on data quality, cable delays below 13 ns (4 m) typically remain uncorrected, which results in less than 145 degrees of phase difference over the 30.72 MHz

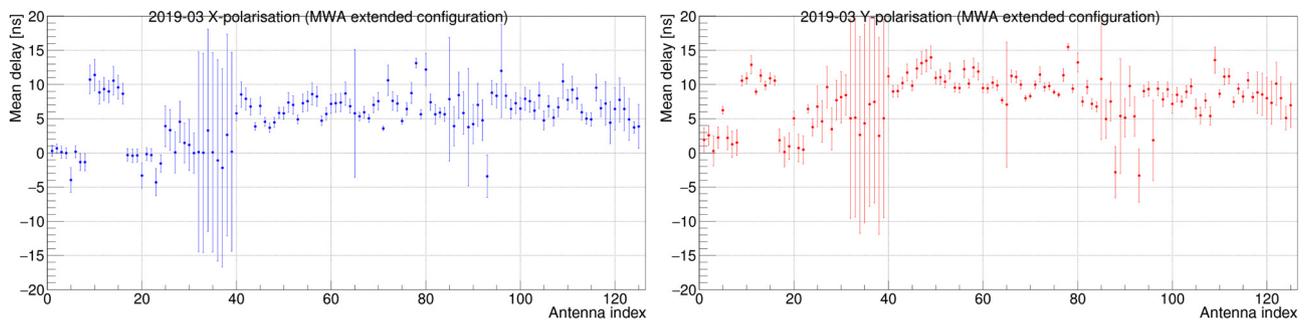


Figure 7. Mean and standard deviation (error bars) of delays for each MWA tile during March 2019 (the MWA in extended configuration). Left image (blue points) is X polarisation and right image (red points) is Y polarisation. These plots are based on data points as shown in Figure 6 for all MWA tiles and are good summary diagnostic plots of the telescope stability. For instance, indexes 9–16 (values around 10 ns) from Tiles 071 to 078 had unaccounted cable length of around 3.23 m, which was not corrected over that month. Large errors in indexes from 32 to 40 (Tiles 121–128) are due to step-like receiver reboots during the month and the same applies to indexes 25–32 (Tiles 111–118). Indexes 65 and 89 (Tile 62 and LBC4, respectively) had cable lengths adjusted in March 2019 corresponding to 18.5 and 5 ns (thus large error bars).

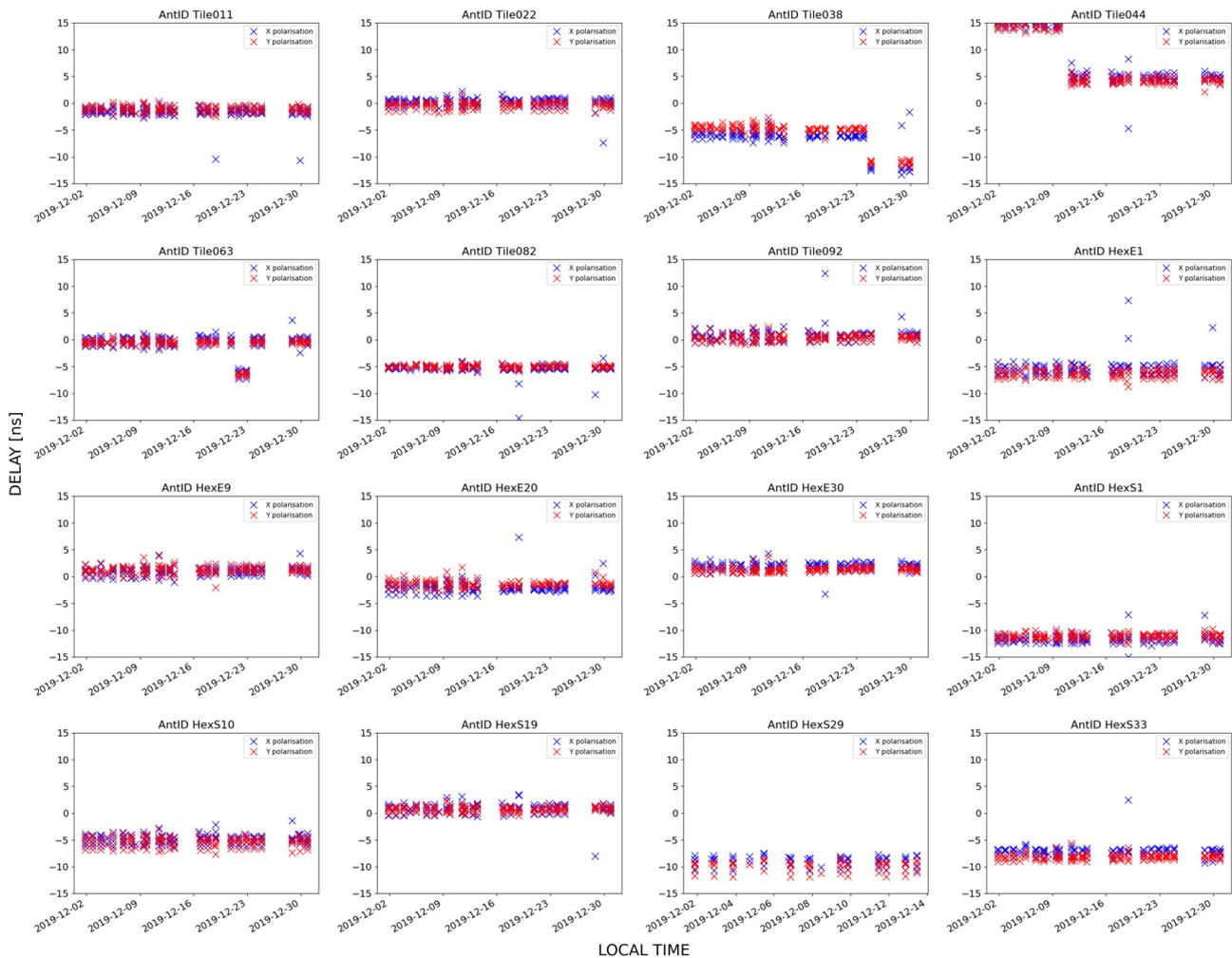


Figure 8. Delay (ns) fitted to the phase of calibration solutions as a function of time in December 2019 (with the MWA in the compact configuration) (Wayth *et al.* 2018). Sixteen tiles representing typically observed behaviour are shown. The short sudden ‘dip’ of about 6 ns observed on Tile 063 was due to a receiver reset (affects all eight tiles connected to a particular receiver). The delay returned to the pre-reset value after a few days when the next reboot occurred. A step of about 10 ns on Tile 044 was due to the adjustment of its cable length in the database. It can be seen that when receivers are not rebooted, the telescope remains very stable over timescales of weeks.

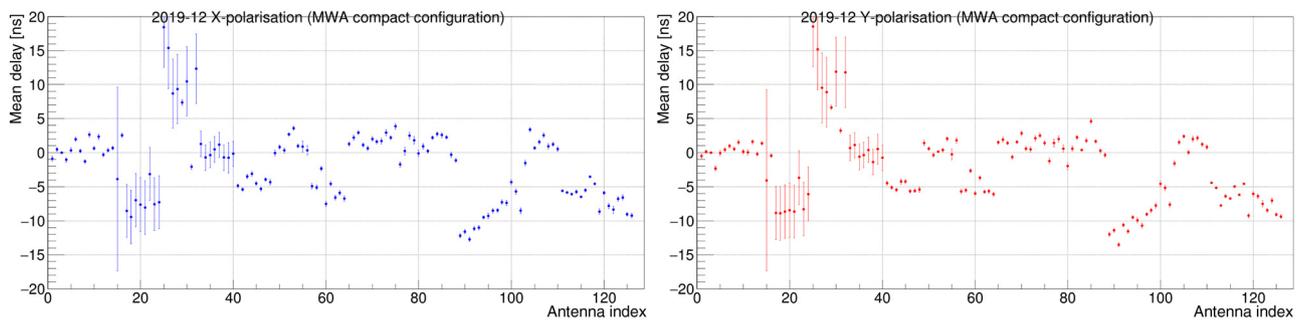


Figure 9. Mean and standard deviation (error bar) of delays for each MWA tile during December 2019 (the MWA in compact configuration). Left image (blue points) is X polarisation and right image (red points) is Y polarisation. These plots are based on data points shown in Figure 8 for all MWA tiles and are good summary diagnostic plots of the telescope stability. For instance, indexes 17–24 (Tiles 031–038) and 25–32 (Tiles 041–048) have large standard deviations (error bars) due to poor receiver reboots which occurred on 2019 December 25 and 2019 December 10, respectively.

band. However, if the delay exceeds 13 ns, the cable length in the instrument set-up database is updated based on the fitted value. Usually, after re-configuration between the compact and extended configurations, several tiles need cable length adjustments in the instrument description database in order to avoid large, uncorrected cable delays (fast ‘phase wraps’) that, if uncorrected, reduce the MWA sensitivity. The calibration system also helped to identify situations when coaxial cables from two tiles were accidentally swapped at a receiver input during a re-configuration between the compact and extended MWA configurations causing large delays (due to cable length from a different tile being used to correct the phase).

The MWA instantaneous observing bandwidth is 30.72 MHz, which is typically placed between 50 and 350 MHz. Thus, we could not perform the fit of a straight line over the full band (starting from zero frequency). Moreover, the combination of sky and beam models used in the calibration are usually not a perfect representation of the sky and instrument. Therefore, we allowed the intercept to be a free parameter of the fit. We verified that the fitted values of the intercept are also very stable over time (excluding times when receivers are rebooted) and they are often close to zero or a multiplicity of 360° . Hence, with the future improvements in the sky model based on the recent extensions of the GaLactic Extragalactic All-sky MWA (GLEAM) catalog (Hurley-Walker et al. 2019), we will consider constraining the intercept value to be either zero or a multiplicity of 360° .

4.2. Providing calibration solutions to the new MWA correlator

The near-real-time pipeline will be used to provide calibration solutions for the new MWA ‘fringe-stopping’ correlator, which is currently in development (Morrison et al. in preparation). The MWA telescope is very stable (Section 4.1) and hence one or two sets of calibrations per 24 h interval should be sufficient. However, if it turns out to be insufficiently accurate, the calibration solutions for the new correlator will be updated more often.

A dedicated calibration server will be deployed at the MRO which will enable immediate direct access to visibility files generated by the MWA correlator. This will significantly speed up the calibration process by eliminating time required to transfer data from the MRO to the PSC archive making the pipeline a truly real-time one.

4.3. Monitoring of calibrator field images for transients and ionospheric quality

For the last 3 yr of the MWA operation, the near-real-time calibration pipeline produced control images of the standard MWA calibrators: Pictor A (537 images); Centaurus A (342 images); Hydra A (323 images); Hercules A (197 images); and 3C444 (214 images). There are even more archival images (before 2016) reduced when populating the MWA ASVO database with the archival data, which opens a possibility of radio-transient searches in these fields over a long-time baseline similar to those performed by Bower & Saul (2011) with the VLA. Roughly, 1/3–1/2 of these images are at the MWA optimal frequency of 154.88 MHz. We have executed the AEGEAN source finderⁿ on these images in order to find transient candidates and catalog sources to a PostgreSQL database. Analysis is ongoing. We are also planning to extend the existing near-real-time pipeline and look for transient candidates on a daily basis. The results of these searches will be reported in a separate publication. Finally, such a database populated soon after the calibrator field data are collected provides an excellent opportunity to calculate the mean offset of the sources from their nominal positions in the GLEAM catalog (Hurley-Walker et al. 2017; Wayth et al. 2015) or other catalogs and provide early information of the given night’s data quality. However, in such a case, the database should be populated more densely with at least one observation every hour (or more if possible) as the ionosphere can change on timescales of hours during the night.

5. Summary

The MWA ASVO calibration component opens a new avenue for researchers worldwide to download calibrated MWA data, create sky images using standard radio astronomy software packages, and analyse these images for multiple purposes. The development of the calibration component of the MWA ASVO interface is a very important contribution to the astronomical community in Australia and beyond, providing access to the MWA data archive to every researcher without requiring a deep knowledge of the instrument. Using the recent sky models obtained from the MWA data (Hurley-Walker et al. 2019; Hurley-Walker et al. 2017), it will be possible to further improve calibration solutions in the database and consequently improve the quality of the resulting images. We

ⁿ<https://github.com/PaulHancock/Aegean>

expect that this endeavour will facilitate greater use of the system by researchers from outside the MWA Collaboration using MWA data.

The development of the CALDB triggered the establishment of an automated data reduction pipeline, which has been used in near-real time for daily monitoring of the quality of the MWA calibration solutions and hence the interferometric performance of the telescope. The pipeline also produces sky images which can be used for monitoring the quality of the ionosphere and looking for transient objects on a daily basis.

Finally, this work has been a starting point to develop a database of calibration solutions for the upcoming low-frequency component of the Square Kilometre Array telescope. Based on the MWA experience, a similar CALDB was developed to store calibration solutions from the SKA-Low prototype stations Aperture Array Verification Systems (AAVS-1 and AAVS-2) and Engineering Development Array 2 (Wayth *et al.* in preparation) already deployed at the MRO. This database will be further extended in order to handle more SKA-Low stations as they will soon be built at the MRO.

Acknowledgements.

People We would like to thank the anonymous referee for the prompt review of our manuscript.

Facilities. This scientific work makes use of the Murchison Radio-astronomy Observatory (MRO), operated by CSIRO. We acknowledge the Wajarri Yamatji people as the traditional owners of the Observatory site.

Funding. Support for the operation of the MWA is provided by the Australian Government (NCRIS), under a contract to Curtin University administered by Astronomy Australia Limited. Development of the MWA ASVO was funded via the Australian Research Data Commons (ARDC), administered by Astronomy Australia Limited. Parts of this research were supported by the Australian Research Council Centre of Excellence for All Sky Astrophysics in 3 Dimensions (ASTRO 3D), through project number CE170100013. This work was further supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia. DLK was supported by NSF grant AST-1816492.

Software. We acknowledge the work and support of the developers of the following following Python packages: Astropy (Astropy Collaboration *et al.* 2013; Price-Whelan *et al.* 2018), Numpy (van der Walt, Colbert, & Varoquaux 2011), Scipy (Virtanen *et al.* 2020), Matplotlib (Hunter 2007), and AegeanTools (Hancock, Trott, & Hurley-Walker 2018). We acknowledge developers of the MWA_Tools library. This research has made use of NASA's Astrophysics Data System.

References

- Astropy Collaboration *et al.*, 2013, *A&A*, **558**, A33
- Bower, G. C., & Saul, D. 2011, *ApJ*, **728**, L14
- Greisen, E. W. 2019, Aips memo series, AIPS FITS File Format. AIPS Memo 117
- Hancock, P. J., & Trott, C. M., Hurley-Walker, N. 2018, *PASA*, **35**, e011
- Hunter, J. D. 2007, *Computing in Science & Engineering*, **9**, 90
- Hurley-Walker, N., *et al.* 2017, *MNRAS*, **464**, 1146
- Hurley-Walker, N., *et al.* 2019, *PASA*, **36**, e047
- Lane, W. M., Cotton, W. D., van Velzen, S., Clarke, T. E., Kassim, N. E., Helmboldt, J. F., Lazio, T. J. W., & Cohen, A. S. 2014, *MNRAS*, **440**, 327
- Line, J. L. B. 2018, PUMA: Low-frequency radio catalog cross-matching (ascl:1807.022)
- McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in *Astronomical Data Analysis Software and Systems XVI*, ed. R. A. Shaw, F. Hill, & D. J. Bell, *Astronomical Society of the Pacific Conference Series*, Vol. 376, 127
- Offringa, A. R., van de Gronde, J. J., & Roerdink, J. B. T. M., 2012, *A&A*, **539**, A95
- Offringa, A. R., *et al.* 2014, *MNRAS*, **444**, 606
- Offringa, A. R., *et al.* 2015, *PASA*, **32**, e008
- Offringa, A. R., *et al.* 2016, *MNRAS*, **458**, 1057
- Price-Whelan, A. M., *et al.* 2018, *AJ*, **156**, 123
- Sault, R. J., Teuben, P. J., & Wright, M. C. H. 1995, *A Retrospective View of MIRIAD*, 433
- Schwarz, G. 1978, *Annals of Statistics*, **6**, 461
- Sokolowski, M., *et al.* 2017, *PASA*, **34**, e062
- Tingay, S. J., *et al.* 2013, *PASA*, **30**, e007
- Tremblay, S. E., *et al.* 2015, *PASA*, **32**, e005
- Virtanen, P., *et al.* 2020, *Nature Methods*, **17**, 261
- Wayth, R. B., *et al.* 2015, *PASA*, **32**, e025
- Wayth, R. B., *et al.* 2018, *PASA*, **35**, 33
- van der Walt, S., Colbert, S. C., & Varoquaux, G. 2011, *Computing in Science and Engineering*, **13**, 22