GPCAL: A Generalized Calibration Pipeline for Instrumental Polarization in VLBI Data

Jongho Park1,2,5, Do-Young Byun3,4, Keiichi Asada1, and Youngjoo Yun3

1 Institute of Astronomy and Astrophysics, Academia Sinica, P.O. Box 23-141, Taipei 10617, Taiwan; jpark@asiaa.sinica.edu.tw
2 Department of Physics and Astronomy, Seoul National University, Gwanak-gu, Seoul 08826, Republic of Korea
3 Korea Astronomy and Space Science Institute, Daejeon-dae-ro 776, Yuseong-gu, Daejeon 34055, Republic of Korea
4 University of Science and Technology, Gajeong-ro 217, Yuseong-gu, Daejeon 34113, Republic of Korea

Received 2020 July 29; revised 2020 November 3; accepted 2020 November 18; published 2021 January 12

Abstract

We present the Generalized Polarization CALibration pipeline (GPCAL), an automated pipeline for instrumental polarization calibration of very long baseline interferometry (VLBI) data. The pipeline is designed to achieve a high calibration accuracy by means of fitting the instrumental polarization model, including the second-order terms, to data from multiple calibrator sources simultaneously. It also allows for using more accurate linear polarization models of calibrators for D-term estimation compared to the conventional way that assumes similar linear polarization and total intensity structures. This assumption has been widely used in the existing packages for instrumental polarization calibration but could be a source of significant uncertainties when there is no suitable calibrator satisfying the assumption. We demonstrate the capabilities of GPCAL by using simulated data, archival Very Long Baseline Array (VLBA) data of many active galactic nucleus (AGN) jets at 15 and 43 GHz, and our Korean VLBI Network (KVN) observations of many AGN jets at 86, 95, 130, and 142 GHz. The pipeline could reproduce the complex linear polarization structures of several sources shown in the previous studies using the same VLBA data. In addition, GPCAL reveals a complex linear polarization structure in the flat-spectrum radio quasar 3C 273 from the KVN data at all four frequencies. These results demonstrate that GPCAL can achieve a high calibration accuracy for various VLBI arrays.

Unified Astronomy Thesaurus concepts: Very long baseline interferometry (1769); Polarimetry (1278); Astronomy software (1855); Astronomy data analysis (1858)

1. Introduction

Very long baseline interferometry (VLBI) is the technique that enables us to achieve very high angular resolution by using widely separated antennas as elements of an interferometer array. Polarization observations with VLBI have revealed the ordering and orientation of magnetic fields in and around nonthermal radio sources, which are believed to play a critical role in their formation and evolution. For example, observations of orientations of the magnetic fields in the jets of various active galactic nuclei (AGNs) suggest the presence of shocks in the jets (e.g., Lister & Homan 2005; Jorstad et al. 2007). In some AGN jets, Faraday rotation measure (RM) gradients transverse to the jets were revealed, which indicates the presence of helical magnetic fields wrapping around those jets (e.g., Asada et al. 2002; Gabuzda et al. 2004; Zavala & Taylor 2005). Also, the magnitude of the RM was found to systematically increase with decreasing distance from the black hole in the jet of the nearby radio galaxy M87 (Park et al. 2019) and increase with frequency in the cores of blazars (e.g., Jorstad et al. 2007; O’Sullivan & Gabuzda 2009; Hovatta et al. 2012, 2019; Kravchenko et al. 2017; Park et al. 2018). These results suggest that AGN jets tend to be in the environments of higher electron densities and/or stronger magnetic fields at shorter distances from the central engines.

An ideal circular polarization feed would respond to only one polarization. However, in reality, any feed will also respond to the other polarization signal, introducing “instrumental” polarization signals to the data. These signals are of the same order as the intrinsic source polarization signals in many cases and must be properly estimated and removed from the data. A task implemented in the Astronomical Image Processing System (AIPS; Greisen 2003) based on the linearized leakage model (Leppanen et al. 1995; see also, e.g., Cotton 1993; Roberts et al. 1994 for more details of instrumental polarization calibration), LPCAL has been widely used for instrumental polarization calibration of VLBI data. It has been very successful for a great deal of studies using various VLBI arrays, such as the Very Long Baseline Array (VLBA; e.g., Jorstad et al. 2017; Lister et al. 2018), Global mm-VLBI Array (GMVA; e.g., Casadio et al. 2017), High Sensitivity Array (HSA; e.g., Hada et al. 2016), RadioAstron space VLBI mission (e.g., Gómez et al. 2016), and Korean VLBI Network (KVN; e.g., Park et al. 2018).

Nevertheless, there are some circumstances in which one needs different calibration strategies and improved calibration accuracy. First of all, for global VLBI arrays such as the Event Horizon Telescope (EHT; Event Horizon Telescope Collaboration et al. 2019a, 2019b, 2019c, 2019d, 2019e, 2019f) and GMVA, a common sky area for a single calibrator for some antennas can be quite limited. In this case, the parallactic angle coverages of the calibrator for those antennas would also be limited, resulting in a relatively inaccurate calibration. Since both the common sky area and parallactic angle coverage are often sensitive to the source’s decl. (see, e.g., Figure 2 in Trippe et al. 2010), using multiple calibrators at different declinations to model the instrumental polarization signals can help to improve the calibration accuracy.

Second, LPCAL relies on the similarity assumption, which assumes that linear polarization structures of calibrators are proportional to their total intensity structures (Cotton 1993; Leppanen et al. 1995). However, this assumption may not...
always hold, especially at high frequencies. Most calibrators for VLBI observations are resolved and show significant variability in their source structures. Thus, it is often challenging to have suitable calibrators satisfying the assumption in the data, even though observers select the calibrators based on the information of the source’s linear polarization structures from previous observations.

Third, some VLBI arrays, such as the KVN and EHT, do not have many antennas (Park et al. 2018; Event Horizon Telescope Collaboration et al. 2019b). It is challenging for those arrays to have instrumental polarization adequately removed from the data due to the small number of measurements\(^7\) (baselines). Combining the results from multiple calibrators could mitigate the difficulty. Still, it is generally not straightforward to take into account different signal-to-noise ratios (S/Ns) and parallactic angle coverages of various sources for combining the results.

Lastly, some heterogeneous VLBI arrays, such as the EHT and GMVA in conjunction with the Atacama Large Millimeter/submillimeter Array (ALMA; e.g., Event Horizon Telescope Collaboration et al. 2019a; Issaoun et al. 2019) and the HSA including the phased-up Very Large Array (VLA; e.g., Hada 2017), have very different sensitivities among different stations. In this case, the fitting would be dominated by the baselines to the most sensitive stations. If the antenna gains of those sensitive stations are not well corrected and there are remaining systematic errors in the data, the fitting solutions of all other stations can be distorted. One can scale the visibility weights of the sensitive stations down for fitting to avoid this problem, as is done for the imaging of the shadow of the supermassive black hole in M87 with CLEAN\(^8\) (Event Horizon Telescope Collaboration et al. 2019d).

These motivated us to develop a new pipeline for instrumental polarization calibration of VLBI data, named the Generalized Polarization CALibration pipeline (GPCAL). It allows us to (i) fit the instrumental polarization model to data from multiple calibrator sources simultaneously, (ii) use more accurate linear polarization models of calibrators for fitting, (iii) flexibly change the visibility weights of each station, and (iv) easily check the fitting results and statistics. It is based on AIPS and Difmap (Shepherd 1997), which have been widely used for calibration and imaging of VLBI data for a long time. We implemented external scripts written in Python in the pipeline only for the impossible or difficult parts to deal with AIPS and Difmap. This makes the pipeline more reliable and friendly to users who are already familiar with that software.

In Section 2, we describe the model of instrumental polarization employed in GPCAL. The general calibration scheme of GPCAL is explained in detail in Section 3. We verify the pipeline and demonstrate the capabilities of GPCAL by using simulated VLBI data and real data observed with different VLBI arrays at different frequencies in Section 4. We present a concluding summary in Section 5.

\(^7\) It would be important for those arrays to keep searching for good calibrators that are either very compact or have very low degrees of polarization, although that appears to be challenging at millimeter wavelengths (e.g., Casadio et al. 2017).

\(^8\) Finding the “sweet spot” of the relative weights between stations may require a dedicated investigation through, e.g., a parameter survey using simulated data (Event Horizon Telescope Collaboration et al. 2019d).

### 2. D-term Model

We follow Leppanen et al. (1995) for description of a model for the interferometer response, which relates the measured cross-correlations \((r_{RR}^*, r_{LL}^*, r_{RL}^*, r_{LR}^*)\) and the true visibilities \((\mathcal{A} \mathcal{R}, \mathcal{L} \mathcal{L}, \mathcal{A} \mathcal{L}, \mathcal{L} \mathcal{R})\) on a baseline \(mn\):

\[
\begin{aligned}
r_{RR}^{mn} &= G_m G_n e^{i \phi_1} r_{RR} + D_m e^{i \phi_2} r_{RL} + D_n e^{i \phi_3} r_{LR} + D_{mn} e^{i \phi_4} r_{LL} \\
&= G_m G_n e^{i \phi_1} r_{RR} + D_m e^{i \phi_2} r_{RL} + D_n e^{i \phi_3} r_{LR} + D_{mn} e^{i \phi_4} r_{LL},
\end{aligned}
\]

where the asterisk denotes a complex conjugate, \(G\) is the complex antenna gain, \(D\) is the leakage factor (so-called “D-term”), and \(\phi\) is the antenna field rotation angle. Subscripts denote antenna numbers, and superscripts denote polarization. The field rotation angle is a function of the elevation angle \((\theta_e)\) and parallactic angle \((\psi_{par})\), depending on antenna mounts,

\[
\phi = f_a \theta_e + f_p \psi_{par} + \phi_{off},
\]

where \(\phi_{off}\) is a constant offset, which is expected when the antenna feed is rotated with respect to the azimuth axis, which is the case for, e.g., ALMA (see Section 4.2 of the ALMA Cycle 7 Technical Handbook). Cassegrain mounts have \(f_p = 1\) and \(f_a = 0\); thus, the field rotation angle is equivalent to the parallactic angle, except for the constant offset. Nasmyth mounts have \(f_p = 1\) and \(f_a = +1\) for Nasmyth-Right type and \(f_a = -1\) for Nasmyth-Left type. The true cross-hand visibilities are related to the Stokes parameters as

\[
\mathcal{A} \mathcal{R} = 2 + j \chi, \quad \mathcal{R} \mathcal{R} = 2 - j \chi = \mathcal{P}^*,
\]

where \(\mathcal{P}\) and \(\mathcal{R}\) are the Fourier transforms of the source’s Stokes \(Q\) and \(U\) on the sky, respectively; \(\mathcal{P}\) is the Fourier transform of the complex polarization \(P \equiv p e^{i \chi}\); \(p\) is the fractional polarization; \(I\) is the total intensity emission on the sky; and \(\chi\) is the electric vector position angle (EVPA; e.g., Roberts et al. 1994). It is assumed by GPCAL that the field rotation angles were already corrected, and the antenna gains were corrected except for the phase offset between right circular polarization (RCP) and left circular polarization (LCP) at the reference antenna.\(^9\) Then, one can write the model cross-hand visibilities \((\tilde{r}_{RR}, \tilde{r}_{LR})\) for each measurement at \((u, v)\)

\[^9\] The phase offsets between polarizations in other stations are expected to be removed during the global fringe fitting and self-calibration (see, e.g., Schwab & Cotton 1983; Roberts et al. 1994; Cotton 1995a, 1995b; Leppanen et al. 1995). A single phase offset \((e^{i \phi_{off}})\), which is believed to originate from the instrumental phase offset between polarizations at the reference antenna, will remain in all baselines. This offset is usually assumed to be constant during the observations and can be corrected after D-term correction (the so-called EVPA calibration). This is absorbed in different terms in Equation (4) such that \(\mathcal{P} \rightarrow \mathcal{P} e^{i \phi_{off}}\), \(D_{RR} \rightarrow D_{RR} e^{i \phi_{off}}\), and \(D_{LL} \rightarrow D_{LL} e^{-i \phi_{off}}\).
coordinates as

$$r_{mn}^{RL}(u, v) = \mathcal{P}(u, v) + D_{m}^{L}e^{2i\phi}r_{mn}^{LL}(u, v) + D_{n}^{L}e^{2i\phi}r_{mn}^{RR}(u, v)$$

and

$$r_{mn}^{LR}(u, v) = \mathcal{P}(u, v) + D_{m}^{R}e^{-2i\phi}r_{mn}^{RR}(u, v) + D_{n}^{R}e^{-2i\phi}r_{mn}^{LL}(u, v).$$

The antenna gains are assumed to be perfectly corrected, and the true parallel-hand visibilities ($\mathcal{P}_{R}, \mathcal{P}_{L}$) are replaced with the measured parallel-hand visibilities ($r_{RR}, r_{LL}$) in these equations. To derive the best-fit D-terms, GPCAL fits these model equations to the measured cross-hand visibilities.

3. GPCAL Calibration Procedures

3.1. Modeling of Source Polarization Structure

Equation (4) requires the antenna field rotation angles and a source polarization model ($\mathcal{P}$) for each visibility data point. The former is a purely geometrical quantity depending on antenna positions and mounts and can be easily computed (e.g., Cotten 1993). However, the latter is difficult to constrain directly and requires some assumptions and strategies, which will be briefly discussed below.

3.1.1. Similarity Assumption

The standard method is to assume that the linearly polarized structures of calibrators are proportional to their total intensity structures, the so-called “similarity” approximation (Cotten 1993). However, this might be an oversimplification for most calibrators, especially at high frequencies, which usually show variations in both fractional polarization and EVPA from region to region (e.g., Jorstad et al. 2007; Lister et al. 2018). One can roughly take into account these variations by splitting the source’s total intensity CLEAN models into several submodels ($I_{s}$) and applying the similarity approximation to each submodel. In other words, each total intensity submodel has a constant fractional polarization and EVPA across the submodel region (Leppanen et al. 1995). This can be expressed as

$$\mathcal{P}(u, v) = \sum_{s} p_{s} \mathcal{F}_{s}(u, v),$$

where $\mathcal{F}_{s}(u, v)$ is the Fourier transform of the submodel $I_{s}$ and $p_{s} = (Q_{s} + jU_{s})/I_{s}$ is the complex coefficient. With this approximation, the number of free parameters in the fitting would be $4N_{\text{int}} + \sum_{\text{cal}} 2N_{\text{cal}}$, where $N_{\text{int}}$ denotes the number of antennas and $N_{\text{cal}}$ is the number of submodels for each calibrator. The coefficient $4$ is from the real and imaginary parts of the D-terms of the RCP and LCP for each antenna, and $\sum_{\text{cal}} N_{\text{cal}}$ is the total number of submodels for all calibrators used in the fitting. The coefficient $2$ comes from the real and imaginary parts of $p_{s}$.

3.1.2. Instrumental Polarization Self-calibration

The similarity assumption may not hold in some cases. This is difficult to predict before observations because many calibrators used for VLBI observations show significant variability in their source structures. A possible solution to be nearly free from the similarity assumption and achieve a better calibration accuracy is as follows.

1. Obtain the best-fit D-terms by using the similarity assumption and remove the D-terms from the data.
2. Produce model visibilities $\mathcal{P}(u, v) = \mathcal{F}(u, v) + j\mathcal{F}(u, v)$ and $\mathcal{P}(u, v) = \mathcal{F}(u, v) - j\mathcal{F}(u, v)$ from imaging of the source’s Stokes $Q$ and $U$ with CLEAN (Högblom 1974) using the D-term corrected data.
3. Fit Equation (4) to the D-term uncorrected data and solve for the D-terms only by using the model visibilities $\mathcal{P}(u, v)$ constrained in step 2. Remove the D-terms from the data using the new best-fit D-term estimates.
4. Iterate steps 2 and 3 until the solutions and fitting statistics are converged.

This scheme is very similar to self-calibration of parallel-hand data, which iterates (i) imaging of the source’s total intensity structures and (ii) solving for antenna gains using the model visibilities. This approach was therefore named “instrumental polarization self-calibration” (Cotton 1995b; see Section 15.4.3). In this case, the number of free parameters in the fitting is $4N_{\text{int}}$ because the source polarization information is separately obtained by CLEAN prior to the fitting (in step 2).

3.2. Calibration Procedures

This subsection describes the calibration procedures employed in GPCAL. The pipeline uses ParselTongue, a Python interface to AIPS (Kettenis et al. 2006), for using relevant AIPS tasks. It reads the input parameters provided by users and runs all of the calibration procedures accordingly. First, self-calibrated UVFITS files and image files containing the CLEAN components of calibrators are loaded into AIPS. If self-calibration was performed in Difmap, which assumes that antenna gains for RCP and LCP are the same, then GPCAL can perform an additional self-calibration using CALIB in AIPS to correct the possible remaining antenna gain offsets between polarizations. If requested, GPCAL splits the total intensity CLEAN models into several submodels using CCDET in AIPS. The submodel split can be done manually by providing ASCII files that contain each submodel’s locations on the map or automatically by CCDET. Then, the model visibility corresponding to each measurement at $(u, v)$ for each submodel, $\mathcal{F}_{s}(u, v)$ in Equation (5), is computed by the AIPS task UVSUB. For each visibility measurement, the antenna field rotation angles are computed by using the source coordinates, antenna positions, and antenna mounts in the headers of the UVFITS files.

Then, GPCAL fits Equation (4) to the observed cross-hand visibility data for each baseband channel (often called an intermediate frequency, IF) using the nonlinear least-squares fitting algorithms implemented in Scipy. The similarity assumption is used at this stage using the CLEAN submodels.

---

10 GPCAL does not assume $RR = LL$, which indicates that it takes into account possible nonzero Stokes $V = (RR - LL)/2$ in the data (Equation (1)). However, the circular polarization of AGN jets is expected to be small ($\lesssim 1\%$); e.g., Wardle et al. 1998; Homan & Wardle 1999; Homan & Lister 2006), and one must correct the antenna gains, especially the gain offsets between RCP and LCP, with good accuracy to obtain the weak source-intrinsic circular polarization signals. The gain offset correction usually requires a careful investigation using many calibrators observed during the same run (e.g., Homan & Wardle 1999; Homan & Lister 2006), and thus, if there are not many calibrators in the data or obtaining the source’s circular polarization is not a primary goal of the study, it is a reasonable approach to perform self-calibration assuming $RR = LL$ (e.g., Jorstad et al. 2005). GPCAL allows users to follow the same strategy.

11 https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html
produced by CCEDT. If multiple calibrators are requested to be used, then it assumes the same D-terms for the calibrators and different source polarization terms for different sources. The fitting algorithm uses all of the visibilities of the requested calibrators simultaneously, using the visibility weights stored in the UVFITS files. Thus, calibrators having higher S/Ns would affect the fitting more, which is a good strategy if systematic uncertainties in the data, such as antenna gains, have been corrected with good accuracy. The GPCAL pipeline loads the UVFITS files of all of the sources specified in the input parameters into AIPS for applying the best-fit D-terms. Additional self-calibration with CALIB can also be performed for these sources if requested. The AIPS antenna tables of the UVFITS files are updated with the best-fit D-terms. The D-term corrected UVFITS files are produced and exported to the working directory by SPLIT and FITTP in AIPS, respectively. The D-terms in both the parallel and cross-hand visibilities, including the second-order terms (Equation (1)), are corrected.

If users ask to perform instrumental polarization self-calibration, then GPCAL executes further calibration procedures. It employs a simple Difmap script, which reads the D-term corrected UVFITS files and the CLEAN windows used for total intensity imaging provided by users. The script performs imaging of the calibrators’ Stokes $Q$ and $U$ with CLEAN until the peak intensity within the windows in the dirty maps reaches the map rms noise times a certain factor specified by users. After the imaging is completed for all of the calibrators for instrumental polarization self-calibration, the images are loaded into AIPS. The Stokes $Q$ and $U$ models for each visibility measurement are extracted by UVSUB, from which $P(u, v)$ in Equation (4) is computed. Again, GPCAL fits the model (Equation (4)) to the visibilities of all of the specified calibrators simultaneously but this time solves for the D-terms only by using the model source polarization visibilities. The best-fit D-terms are applied to the data with SPLIT, and the D-term corrected UVFITS files are exported to the working directory with FITTP.

This procedure, obtaining the model polarization visibilities with CLEAN, solving for the D-terms using the model visibilities and the D-term uncorrected data, and producing new D-term corrected UVFITS files, is repeated as many times as specified in the input parameters. The calibrators for this procedure do not have to be the same as those used for the initial D-term estimation using the similarity assumption. Calibrators with high fractional linear polarization with complex polarization structures can be usable. Nevertheless, selecting good calibrators having either very low degrees of linear polarization or compact linear polarization structures is important, especially for the initial D-term estimation. This is because instrumental polarization self-calibration, very similar to total intensity imaging and self-calibration, would work well only when the initial D-term estimates are reasonably close to the real D-terms. Similarly, one should avoid using calibrators having poor field rotation angle coverage for many stations and low S/Ns. Those calibrators would easily degrade the D-term estimates. We present the results of a simple test that demonstrates the importance of selecting good calibrators in Appendix A.

Several kinds of figures are produced by GPCAL, including plots showing (1) the antenna field rotation angles of the calibrators, (2) the Stokes $Q$ and $U$ visibilities (amplitudes and phases) of the calibrators and their best-fit models, and (3) the fitting residuals in units of visibility errors for each station, averaged over each scan and all baselines. The first plots allow users to check whether the field rotation angles of the calibrators have wide enough coverage for all antennas, which is essential for robust D-term estimation. The second plots are useful for examining the overall goodness of fit for each baseline. The third plots help identify some problematic scans for some stations, showing large fitting residuals. Large residuals could be caused by imperfect antenna gain correction or the elevation dependence of the D-terms of some stations, which violates the assumption of constant D-terms during observations used in the model equation.

The GPCAL pipeline computes the reduced $\chi^2$ of the fit ($\chi^2_{\text{red}}$) for each step (using the similarity assumption and the rth iteration of instrumental polarization self-calibration) and produces a plot of $\chi^2_{\text{red}}$ as a function of the steps. It also produces plots of the fitted D-terms on the real and imaginary plane for each IF. The fitted D-terms are saved in ASCII files for each step. A log file containing all of the procedures in AIPS, Difmap, and GPCAL is also produced. Users can investigate the cause of potential errors during the pipeline run with the log file. An overview of the pipeline procedures is summarized in Figure 1. The pipeline is publicly available at https://github.com/jhparkastro/gpcal.

3.3. Optional Functions

The GPCAL pipeline provides functions that could be useful in some specific circumstances, which are briefly summarized as follows.

---

**Figure 1.** Overview of the GPCAL procedures. The pipeline largely consists of two stages: (i) the default procedures using the similarity assumption (magenta) and (ii) the optional procedures using instrumental polarization self-calibration (blue). The second stage is iterated as many times as requested by users. The pipeline can use multiple calibrators simultaneously for both stages, and different sets of calibrators can be used for different stages. It uses ParselTongue (Kettenis et al. 2006) for running necessary AIPS tasks (green text in parentheses) and Difmap (Shepherd 1997) for obtaining Stokes $Q$ and $U$ CLEAN models for instrumental polarization self-calibration.
1. It allows one to scale the visibility weights of specific antennas up or down by a constant factor. This is particularly useful for arrays having very different sensitivities among antennas, such as the EHT+ALMA (Event Horizon Telescope Collaboration et al. 2019a), the GMVA+ALMA (Issaoun et al. 2019), and the HSA including the phased-up VLA (Hada 2017). In this case, fitting can be dominated by the most sensitive stations, and the possible residual systematic errors in those stations can distort the fitting solutions for other stations. One can mitigate this effect by down-weighting those station visibilities for fitting (see, e.g., Event Horizon Telescope Collaboration et al. 2019d), and this option is implemented in GPCAL.

2. In some cases, the D-terms of some stations can be constrained externally. Users can use that prior knowledge and fix the D-terms of those stations with the known values for fitting.

3. Some VLBI arrays have very short baselines, e.g., the phased-up ALMA and the Atacama Pathfinder Experiment (APEX) telescope in Chile in the EHT, which provides a baseline length of 2.6 km (Event Horizon Telescope Collaboration et al. 2019c). Calibrators with compact source geometries will be seen as pointlike sources on these short baselines. Therefore, one can estimate the D-terms of the stations comprising the short baselines by assuming pointlike sources for the source polarization terms in the D-term model, i.e., constant $\mathcal{J}(u, v)$ in Equation (4). Since the model is much simpler and has a much smaller number of free parameters compared to using the whole array at one time, the D-terms of those stations can be robustly constrained. The GPCAL pipeline allows for first obtaining the D-term solutions of those stations using the short baselines and then fixing them in the fitting for the rest of the arrays using the whole baselines. One can use multiple baselines and sources simultaneously for the fitting with short baselines to obtain more accurate D-term solutions.

4. Verification and Results

In this section, we evaluate the performance of the pipeline using simulated and real data sets.

4.1. Simulated Data

We used PolSimulate\textsuperscript{12} in the Common Astronomy Software Application (CASA; McMullin et al. 2007) package to produce simulated data. We assumed (i) an array configuration of the VLBA, (ii) an observing frequency of 15.256 GHz with a bandwidth of 64 MHz, (iii) an on-source time of 1 hr spread over 10 scans over 10 hr, (iv) a minimum elevation cutoff of $10^\circ$, (v) an atmospheric opacity at the zenith of 0.05, and (vi) sky, ground, and receiver temperatures of 70, 90, and 50 K, respectively. The D-terms from a few to about 7% were assumed, which are the typical amplitudes of the D-terms seen in the VLBA data (e.g., Attridge et al. 2005; Hada et al. 2016; Jorstad et al. 2017; Lister et al. 2018; Park et al. 2019). No antenna gain error was introduced for the simulation. We simulated three data sets for the source coordinates of OJ 287, 3C 273, and BL Lac, which have been observed by many VLBI arrays and cover a wide range of right ascension. The source geometries are assumed to consist of a few point sources for Stokes $I$, $Q$, and $U$. We considered two different cases for the source structures: (i) the locations of the Stokes $Q$ and $U$ models being coincident with those of the Stokes $I$ models ($P \propto I$) and (ii) the Stokes $Q$ and $U$ models being shifted from the Stokes $I$ models by $\approx 60\%-120\%$ of the FWHMs of the synthesized beams depending on sources ($P \propto I$). The former is an ideal case for using the similarity assumption, while the assumption does not hold at all for the latter case. In Figure 2, we present the example linear polarization maps of the simulated data for a 3C 273–like uv coverage.

We performed imaging of the simulated data sets with CLEAN in Difmap and ran GPCAL using the similarity assumption for the former case ($P \propto I$). We divided the total intensity source models into several submodels in such a way that each knot-like structure is regarded as a submodel. We fitted the D-term model to the data of all three sources simultaneously to verify the multsource fitting capability of GPCAL. Since no antenna gain error was introduced in the simulation, one can expect to reconstruct the assumed ground-truth D-terms nearly perfectly. Some expected sources of minor deviation from the truth values are (i) deconvolution errors in CLEAN, (ii) parallel-hand visibilities distorted by the D-terms, and (iii) thermal noise in the data. The second source comes from the fact that our model (Equation (4)) assumes that the measured parallel-hand visibilities ($r^{LL}$, $r^{RR}$) are the same as the true visibilities ($I^R$, $I^L$). However, the measured visibilities in the simulated data are slightly distorted by the D-terms and not identical to the true visibilities in reality (Equation (1)).

Figure 3 shows the results for the $P \propto I$ case (Figure 2, left). The reconstructed D-terms are consistent with the ground-truth D-terms with an rms error (RMSE) of $\approx 0.06\%$. The reconstructed source polarization terms (the fractional polarizations and EVPAs) are also in good agreement with the assumed source polarizations in the simulation. We present the RMSEs obtained by using the simulated data for only one source and two sources in Table 1. The RMSE becomes smaller when we use more sources, as expected. These results verify that GPCAL can derive the D-terms from multiple calibrators data simultaneously when the linear polarization structures are similar to their total intensity structures.

For the $P \propto I$ case (Figure 2, right), we repeated the above procedures, on top of which we performed instrumental polarization self-calibration with 10 iterations. Figure 4 shows that the reconstructed D-terms using the similarity assumption (iteration zero) significantly deviate from the ground-truth values with an RMSE of $\approx 5\%$. This is because the assumed source geometries are far from the similarity assumption. However, the reconstruction is progressively improved as we iterate instrumental polarization self-calibration and becomes nearly converged to the ground-truth values with an RMSE of $\approx 0.1\%$ after 10 iterations. This result demonstrates that GPCAL can reconstruct the D-terms even from calibrators having complex linear polarization structures, which has been challenging for existing packages like LPCAL.

4.2. Real Data

In this subsection, we apply GPCAL to several real data observed with different VLBI arrays and at different frequencies to evaluate its performance.

\textsuperscript{12} https://launchpad.net/casa-poltools
4.2.1. MOJAVE Data at 15 GHz

We analyzed one of the Monitoring Of Jets in Active galactic nuclei with VLBA Experiments (MOJAVE; Lister et al. 2018) data sets, which have observed many AGN jets with the VLBA at 15 GHz for decades. We selected the observation of 30 sources on 2017 January 28 at a recording rate of 2 Gbps (project code: BL229AE). Since the publicly available data on the MOJAVE...
Three RMSE values obtained by running GPCAL using one source (3C 273, OJ 287, and BL Lac individually) and two sources (three possible combinations of the three sources) are averaged. The RMSE value in the last column is obtained by using all three sources together for fitting.

database are already fully calibrated, including D-term correction, we analyzed the raw data in the VLBA archive. We performed a standard data postcorrelation process with AIPS following Park et al. (2019) and hybrid imaging with CLEAN and self-calibration in Difmap.

We ran the pipeline on the self-calibrated data. We let GPCAL perform additional self-calibration with CALIB in AIPS to correct the potentially remaining gain offsets between the polarizations. We selected five calibrators, 0256+075, 0851+202, 2200+420, 2201+315, and 2258–022, that are bright (total flux greater than 0.5 Jy) and have core-dominated linear polarization structures in this particular epoch for the initial D-term estimation using the similarity assumption. These sources consist of several knot-like structures, which were taken into account for splitting their total intensity CLEAN components into several submodels. The pipeline performed additional instrumental polarization self-calibration with 10 iterations by including six more calibrators, 0300+470, 0415+379, 0430+052, 0502+049, 1151+408, and 1637+574, that are bright but have relatively complex linear polarization structures.

We present an example of the fitting results with the Los Alamos (LA) and Pie Town (PT) baseline for Stokes U data in Figure 5. The model derived by GPCAL, obtained by using 11 sources simultaneously, fits the visibilities for all different calibrators quite well. This is the case for many calibrators having complex polarization structures as well, thanks to the instrumental polarization self-calibration. We also present the contributions from the source polarization terms, i.e., the first terms in Equation (4), and the instrumental polarization terms, i.e., the rest terms in Equation (4). The former shows smooth variations with time, as expected from the smooth changes of (u, v) over time, while the latter varies rapidly with time due to the changes of antenna parallactic angles. In Figure 6, the fitting residuals in units of visibility errors, i.e., |V − ˆV|/σ, where V is the visibility data, ˆV is the model visibility, and σ is the visibility error, are shown for the Saint Croix (SC) station. The residuals are averaged over each scan and all baselines to the station. If the data are represented well by the model, one could expect residuals of ≈1.4σ, which corresponds to the reduced χ² of 1. Therefore, this plot helps to identify some problematic scans for specific stations. The most likely causes of a bad fit for some scans are imperfect antenna gain correction or variable D-terms during the observations. The former effect would be more easily seen in very bright calibrators because their systematic errors usually dominate the thermal noise (the error bars). In other words, the fitting would look good for faint calibrators even if there were moderate residual antenna gains in the data because of the large error bars. The latter would usually appear on scans at very low or high elevations.

Figure 7 shows the reduced χ² of the fits for the different steps of calibration. The statistics are significantly improved between using the similarity assumption (iteration zero) and the first iteration of instrumental polarization self-calibration (iteration one). They are gradually improved with more iterations and become saturated after four or five iterations. This result demonstrates that the similarity assumption does not perfectly hold even for the calibrators having core-dominated linear polarization structures for these data.

After the D-term correction, we corrected the remaining RCP and LCP phase offset at the reference antenna for each IF by comparing the integrated EVPA of 0851+202 (OJ 287) with

---

13. The LA–PT baseline is the shortest baseline of the VLBA data, and the SC station comprises the longest baselines; they are selected as examples.
14. The reduced χ² are slightly larger for IF 6–8 than the other IFs for the particular MOJAVE data. We found that the best-fit D-terms of several antennas become larger at higher frequencies; this trend could also be seen in the cross-power spectra for the cross-hand visibilities. If the D-terms are large, systematic errors such as the time-dependent D-terms could also be more severe (the D-terms can easily change by a few percent during the observations if the D-terms are several tens of percent). This can result in the poor fits for those IFs.
that in the MOJAVE database.\(^{15}\) We produce linear polarization maps of \(1226\text{+}023\text{ (3C 273)}\) and \(0430\text{+}052\text{ (3C 120)}\), which are known to have very complex total intensity and linear polarization structures (e.g., Gómez et al.\(^{2000}\); Asada et al.\(^{2002}\)), and compare them with the MOJAVE maps in the left panels of Figure 8. We found that the distributions of linearly polarized intensity and EVPA are very consistent between the GPCAL and MOJAVE results for both sources. In the right panels of Figure 8, we also present the fractional polarizations and EVPAs at several locations of the jets indicated in the maps. Both quantities are in good agreement between the maps. The MOJAVE program has obtained D-terms with a high accuracy by combining the LPCAL results of many individual sources showing similar D-terms observed in the same run (e.g., Lister & Homan\(^{2005}\); Hovatta et al.\(^{2012}\)). The fact that GPCAL could reproduce nearly identical linear polarization maps to the MOJAVE results demonstrates its capability of achieving a high degree of accuracy in D-term estimation. Normal VLBI programs usually do not observe such a number of sources, and GPCAL will be especially useful in those cases where many calibrators suitable for LPCAL are not available.

It may not be straightforward for users to decide that they should request instrumental polarization self-calibration. In that case, one can check the reduced \(\chi^2\) plot as shown in Figure 7 and make a decision based on whether the statistics improve with instrumental polarization self-calibration or not. Also, if there are calibrators having complex linear polarization structures, it may be questionable whether including those calibrators for instrumental polarization self-calibration would improve or degrade the D-term solutions. We perform a simple test to address this question using simulated data in Appendix B. The result suggests that adding more calibrators can improve the D-term solutions even though they have complex polarization structures. However, this result is based on the simulated data having simple source structures and no antenna gain errors in the data. As one can naturally imagine, the D-term solutions may be degraded if calibrators having poor field rotation angle coverage, low S/Ns, and antenna gain errors that are not well corrected are included for calibration. It is recommended for users to try different combinations of calibrators for the initial D-term estimation and instrumental polarization self-calibration; check the

\(^{15}\) http://www.physics.purdue.edu/astro/MOJAVE/sourcepages/0851+202.shtml

Figure 5. The GPCAL fitting results on the MOJAVE data observed on 2017 January 28–29 for an example baseline of LA–PT for Stokes \(U\) and IF 1 data. The Stokes \(U\) data’s amplitudes and phases for six out of 11 calibrators are shown in different panels. The data are averaged over each scan for better visibility, and the error bars represent the scatter of the data within each scan. The best-fit model (gray solid lines), which consists of the source polarization terms (gray dashed lines) and the instrumental polarization terms (gray dotted lines), is shown. The source polarization terms are obtained from CLEAN of Stokes \(Q\) and \(U\) data for each source during instrumental polarization self-calibration, while the instrumental polarization terms are derived by fitting the model that assumes the same D-terms for different sources (see text for more details). Although the results are shown for each source separately, the model was fitted to the data of all sources simultaneously. The source polarization terms show smooth variations with time, while the instrumental polarization terms vary rapidly with time due to the change of antenna field rotation angles.
We evaluate the performance of GPCAL by using other VLBA data observed as part of the VLBA-BU-BLAZAR (BU) monitoring program at 43 GHz\textsuperscript{16} (Jorstad et al. 2017). We analyzed the data observed on 2015 July 2 (project code: BM413I) and obtained CLEAN images and self-calibrated data, similar to the MOJAVE data analysis. We ran GPCAL using six bright and compact calibrators, 0235+164, 0420–014, OJ 287, 1156–295, 1510–089, and 1749–096, for both the initial D-term estimation using the similarity assumption and additional instrumental polarization self-calibration with 10 iterations. The reduced $\chi^2$ of 1.7–1.9 were obtained for different IFs. The EVPA calibration was done by referring to the integrated EVPA of OJ 287 in the BU database.

In Figure 9, we compare the linear polarization maps of two sources, 3C 84 and Mrk 501, also known for complex linear polarization structures (e.g., Marscher 2016; Nagai et al. 2017), between the GPCAL and BU results. The distributions of linear polarization intensity and EVPA in the jets are consistent between the maps, even for the very weak polarization far from the cores. The consistency can also be seen in the right panels showing the fractional polarizations and EVPAs at several locations of the jets. This result demonstrates the capability of GPCAL for achieving a high D-term estimation accuracy.

4.2.3. KVN Data at 86–142 GHz

Because of its small number (three) of antennas, GPCAL is almost ideally suited for instrumental polarization calibration for the KVN. The limited number of baselines makes calibration quite sensitive to residual systematic errors in the data, and the D-terms from different calibrators often show large dispersion ($\gtrsim 1\%–3\%$; Park et al. 2018). In this case, using many calibrators in a single observing run with GPCAL would help to improve the D-term estimation accuracy. This is because the number of measurements at different antenna field rotation angles increases roughly by a factor of the number of calibrators, while the number of free parameters in the fitting increases by a factor of less than a few in the case of using the similarity assumption and does not increase at all when using instrumental polarization self-calibration.

We observed 11 AGN jets on 2018 February 20–21 with the KVN at 86, 95, 130, and 142 GHz and applied GPCAL to these data. One of the reasons for selecting these data is to verify the performance of the KVN polarimetry at very high frequencies up to 142 GHz, which is a unique advantage of the KVN, with assistance from GPCAL for an accurate D-term estimation. The observations at 86/130 and 95/142 GHz were performed on the first (February 20) and second (February 21) days, respectively. The data were taken at a recording rate of 8 Gbps and divided into a single IF for each frequency. More detailed descriptions of the data analysis and results will be presented in a forthcoming paper (J. Park et al. 2020, in preparation). We performed data reduction, imaging, self-calibration,\textsuperscript{17} and instrumental polarization calibration with GPCAL in a similar manner to the VLBA data analysis (Sections 4.2.1 and 4.2.2). We used four to seven calibrators, which are bright and weakly polarized or moderately polarized with relatively simple structures, for the initial D-term estimation.

\textsuperscript{16} https://www.bu.edu/blazars/VLBAproject.html

\textsuperscript{17} Self-calibration is performed for visibility phases only because there are three antennas.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure6.png}
\caption{The GPCAL fitting residuals of the MOJAVE data for an example station. The fitting residuals in units of visibility errors, $|V−V|/\sigma$, of the Stokes $Q$ (upper panel) and $U$ (lower panel) data are averaged over each scan and all baselines to the SC station. If the data are represented well by the model, then residuals of $\approx 1.4\sigma$ are expected, which corresponds to the reduced $\chi^2$ of 1.}
\end{figure}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure7.png}
\caption{Reduced $\chi^2$ of the fitting to the MOJAVE data with GPCAL. Iteration on the x-axis shows different stages of calibration in the pipeline; iteration zero means an initial D-term estimation using the similarity assumption. Positive iterations mean different iterations of instrumental polarization self-calibration. The results for different IFs are shown with different colors and symbols. The goodness of fit is significantly improved between iterations zero and one, which indicates that the similarity assumption does not perfectly hold for the calibrators used for the initial D-term estimation, even though they have simple and core-dominated linear polarization structures. It is gradually improved with more iterations of instrumental polarization self-calibration and becomes saturated after four or five iterations.}
\end{figure}
estimation using the similarity assumption. We included a few more calibrators having relatively complex polarization structures for instrumental polarization self-calibration with 10 iterations. We obtained the reduced $\chi^2$ of 1.3–2.1 for different bands. The relatively bad goodness of fit for some bands may be understandable, as we could not perform amplitude self-calibration because of the lack of antennas. Thus, there could be nonnegligible gain errors affecting the fitting.

We performed the EVPA calibration by comparing the integrated EVPAs of 3C 279 and OJ 287 with their EVPAs obtained by contemporaneous KVN single-dish observations, as described in Park et al. (2018). We found that 3C 273, among our targets, would be a good source to test the performance of GPCAL. This flat-spectrum radio quasar shows an unpolarized core and moderately polarized knots at $\approx 0.3$ and $\approx 0.7$ mas from the core in the contemporaneous high-resolution BU observation on 2018 February 17 at 43 GHz. This structure would be difficult to obtain if there are significant D-term residuals in the data because the residuals tend to appear as artificial polarization signals in proportion to the total intensity emission (Leppanen et al. 1995).

In Figure 10, we present the linear polarization maps of 3C 273 at four KVN frequencies. We also include the BU polarization map as a reference. All maps are convolved with the synthesized beam at 86 GHz for a proper comparison. We ignored a possible core shift between frequencies for image alignment because the expected core shift between 43 and 130 GHz is $\approx 0.05$ mas (Lisakov et al. 2017), which is much smaller than the convolving beam size. We found that the peak polarization positions are shifted from the cores at all four frequencies, although the large beam size of the KVN does not allow us to see the fine structures that were observed in the high-resolution BU map. Interestingly, the shifts at 86–95 GHz are larger than those at 130–142 GHz.

One of the possible explanations for the different shifts is less depolarization of the jet at higher frequencies (e.g., Sokoloff et al. 1998). If the linear polarization intensity of the inner knot at $\approx 0.3$ mas seen in the BU map becomes larger or the core polarization starts to be detected at higher frequencies due to less depolarization, then the positions of the peak polarization intensity could be shifted toward the core at higher frequencies. The higher degrees of linear polarization at

---

**Figure 8.** Linear polarization maps (left) and polarization properties at several locations marked in the maps (right) of 1226+023 (3C 273; upper panels) and 0430+052 (3C 120; lower panels). The results obtained by GPCAL and taken from the MOJAVE database are shown for comparison. Contours start at 1.7 (0.52) and 3.9 (0.69) mJy beam$^{-1}$ for the left and right maps of 3C 273 (3C 120), respectively, which are three times the off-source image rms noise of each map, and increase by factors of 2. Both the distributions of linearly polarized intensity and EVPA for both sources are consistent between the two results. The fractional polarizations ($m_L$) and the EVPAs ($\chi$) at several locations of the jets are also in good agreement.

---

18 https://www.bu.edu/blazars/VLBA_GLAST/3c273/3C273feb18_map.jpg
≈0.7 mas (marked as “A” in the map) at 86–142 GHz than at 43 GHz may support this interpretation, although it is difficult to investigate the exact origin of the different shifts with the low-resolution maps only. Our results of the 3C 273 polarization using the KVN, showing a misalignment between the linearly polarized intensity peak and total intensity peak positions, demonstrate that GPCAL is capable of achieving a high D-term estimation accuracy and will be useful for future polarimetric studies using the KVN.

5. Summary and Conclusion

We have presented GPCAL, an automated pipeline for instrumental polarization calibration of VLBI data based on AIPS and Difmap. The general calibration procedure of the pipeline follows LPCAL in AIPS, which has been successful for a multitude of studies using various VLBI arrays for a long time. The GPCAL pipeline provides several new functions that can enhance the D-term estimation accuracy.

First, it can fit the D-term model to data from multiple calibrator sources simultaneously. This means that GPCAL properly considers the visibility weights of various sources, which should provide statistically more robust results compared to taking averages of the D-terms from individual sources estimated by LPCAL. Second, it allows for using more accurate linear polarization models of calibrators for D-term estimation than the conventional way using the similarity assumption.
which assumes that the linear polarization structures are proportional to the total intensity structures. This assumption may not hold in many cases, especially at high frequencies, and could be a source of significant uncertainties in the estimated D-terms. Third, it includes the second-order terms in the model and can deal with the case of large D-terms and high source fractional polarization. Lastly, it provides many useful functions, such as (i) changing the visibility weights of some stations for fitting, (ii) fixing the D-terms of some stations to be certain values when external constraints on those D-terms are available, (iii) estimating the D-terms of the stations comprising very short baselines and using them for fitting for the rest of the array, and (iv) providing plots showing the fitting results and statistics, which are useful for identifying some problematic scans or stations.

We have illustrated the capabilities of GPCAL by employing the simulated data and the real data sets observed with different VLBI arrays and at different frequencies. We produced the data simulated with PolSimulate in CASA, assuming simple source geometries consisting of several point sources for Stokes $I, Q, and U$; a VLBA-like array; and uv coverages of three sources. We assumed two cases for the source geometries: one with the locations of the total intensity models being coincident with those of the linearly polarized intensity models ($P \propto I$) and the other with the locations being significantly shifted from each other ($P \propto \chi^2$). We show that GPCAL can reproduce the ground-truth $P \propto \chi^2$-law terms assumed in the simulation very well for both cases by using the data of three sources simultaneously. The latter case was difficult to model in the conventional way using the similarity assumption but could be successfully modeled thanks to the instrumental polarization self-calibration mode implemented in GPCAL.

We have applied GPCAL to the data of the monitoring programs of AGN jets with the VLBA at 15 GHz (the MOJAVE program) and 43 GHz (the VLBA-BU-BLAZAR program). We have shown that GPCAL can fit the model to the data of multiple calibrators simultaneously. The sources having complex linear polarization structures could also be used by performing instrumental polarization self-calibration. This result demonstrates that GPCAL will be very useful when there are no or few calibrators suitable for applying the similarity assumption in the data. We have compared the linear polarization maps of the sources showing complex polarization features obtained by GPCAL and taken from the monitoring program databases. The results are very consistent. These programs have achieved a high D-term estimation accuracy thanks to many good calibrators satisfying the similarity assumption well in their programs. The fact that GPCAL could reproduce nearly identical results to those programs demonstrates its capability of achieving a high D-term estimation accuracy. It will be especially useful for normal VLBI programs, for which it is difficult to have many good calibrators.

We have also applied GPCAL to the data of many AGN jets observed with the KVN at 86, 95, 130, and 142 GHz. Accurate D-term calibration for the KVN is understandably challenging because of the small number of antennas. The pipeline is well-suited to this type of data as well because one can increase the number of measurements by using many calibrators, while the number of free parameters is slightly increased or does not increase. We have shown that the linear polarization maps of 3C 273 at different frequencies obtained by GPCAL successfully recover its complex polarization structure seen in the contemporaneous high-resolution VLBA image at 43 GHz, namely, the unpolarized core and the moderately polarized knots downstream of the core. The images even showed an indication of different amounts of depolarization at different frequencies, although a detailed interpretation is challenging because of the large synthesized beam of the KVN. This result demonstrates that GPCAL would be very useful for instrumental polarization calibration of VLBI arrays not having many antennas.

We note that the pipeline’s current implementation does not take into account possible residual complex antenna gains and time-dependent D-terms for fitting. Also, GPCAL assumes that all calibrators are on equal footing, while in reality, some calibrators should be better than others, depending on field rotation angle coverage, S/Ns, and so on, and one should put...
more weight on the good calibrators for fitting. These will be considered for future developments to further enhance the polarization calibration accuracy. We conclude with a remark on the importance of careful planning of observations. Although GPCAL provides useful functions to overcome the limitations of the existing calibration packages and enhance the calibration accuracy, it is always important to have as many “good” calibrators as possible in the observations. This is especially important for the initial D-term estimation using the similarity assumption, which can affect the D-term estimates in the next steps and the final D-term estimates (Appendix A).

We thank the anonymous referee for detailed comments that significantly improved the manuscript. J.P. acknowledges financial support from the Korean National Research Foundation (NRF) via Global PhD Fellowship Grant 2014H1A2A1018695. J.P. is supported by an EACOA Fellowship awarded by the East Asia Core Observatories Association, which consists of the Academia Sinica Institute of Astronomy and Astrophysics, the National Astronomical Observatory of Japan, the Center for Astronomical Mega-Science, the Chinese Academy of Sciences, and the Korea Astronomy and Space Science Institute. This work is supported by Ministry of Science and Technology of Taiwan grants MOST 109-2112-M-001-025 and MOST 108-2112-M-001-051 (K.A.). This research has made use of data from the MOJAVE database, which is maintained by the MOJAVE team (Lister et al. 2018). This study makes use of 43 GHz VLBA data from the VLBA-BU Blazar Monitoring Program (VLBA-BU-BLAZAR; http://www.bu.edu/blazars/VLBAproject.html), funded by NASA through the Fermi Guest Investigator Program. The VLBA is an instrument of the National Radio Astronomy Observatory. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated by Associated Universities, Inc. We are grateful to the staff of the KVN who helped to operate the array and correlate the data. The KVN and a high-performance computing cluster are facilities operated by the Korea Astronomy and Space Science Institute (KASI). The KVN observations and correlations are supported through the high-speed network connections among the KVN sites provided by the Korea Research Environment Open NETwork (KREONET), which is managed and operated by the Korea Institute of Science and Technology Information (KISTI). Facilities: VLBA(NRAO), KVN(KASI).

Software: AIPS (Greisen 2003), Difmap (Shepherd 1997), ParselTongue (Kettenis et al. 2006), CASA (McMullin et al. 2007), NumPy (van der Walt et al. 2011), Scipy (Virtanen et al. 2020), Pandas (McKinney 2010), Astropy (Astropy Collaboration et al. 2013, 2018), Matplotlib (Hunter 2007).

### Appendix A

#### Importance of Selecting Good Calibrators

We have presented that GPCAL can take into account complex linear polarization structures of calibrators by using instrumental polarization self-calibration (Section 3.1.2). However, this procedure is based on the initial D-term estimation using the similarity assumption. Thus, the initial D-term estimates need to be reasonably accurate to achieve a high accuracy for the final D-term estimates. We performed a simple test to demonstrate the importance of selecting good calibrators for the initial D-term estimation. We obtained D-term estimates for the BU 43 GHz data by using two sources individually, 3C 279 and 1156+295, in the same manner as in Section 4.2.2. Object 3C 279 is known for having a high (∼10%) fractional polarization (e.g., Park et al. 2018) and a complex jet structure at millimeter wavelengths (e.g., Jorstad et al. 2017; Kim et al. 2020) and thus is not a good calibrator for the initial D-term estimation. On the other hand, 1156+295 has a core-dominated linear polarization structure with a fractional polarization of a few percent (Jorstad et al. 2017), which can serve as a good calibrator for the initial D-term estimation.

In Figure 11, we compare the D-terms obtained with those individual sources with our best D-term estimates using multiple sources simultaneously presented in Section 4.2.2. The initial D-term estimates obtained with 3C 279 deviate a lot from the best estimates with an RMSE of ∼1.7%, while those with 1156+295 show a good consistency with an RMSE of ∼0.4%, as expected. The 3C 279 D-terms do improve with more iterations of instrumental polarization self-calibration, reaching an RMSE of ∼0.7% after 10 iterations. The 1156+295 D-terms do not change much with instrumental polarization self-calibration, indicating

**Figure 11.** Comparison of D-terms for the BU 43 GHz data obtained with GPCAL by using individual sources, 3C 279 (blue) and 1156+295 (green), on the y-axis, with the best D-term estimates obtained by using multiple sources simultaneously presented in Section 4.2.2 on the x-axis. The initial D-term estimates obtained by using the similarity assumption (left) and the estimates after one (middle) and 10 (right) iterations of instrumental polarization self-calibration for the individual source estimates are shown, while the same best D-term estimates obtained after 10 iterations of instrumental polarization self-calibration are shown in all three panels. Filled and open symbols indicate the real and imaginary parts of the D-terms, respectively. Circles and squares denote D-terms of RCP and LCP, respectively. The black dashed lines are one-to-one lines between the x- and y-axes.
that the similarity assumption does work well for this source. The final D-term estimates obtained with 3C 279 deviate more from the best estimates than do those with 1156+295, which demonstrates the importance of selecting good calibrators having either very low degrees of polarization or simple linear polarization structures for the initial D-term estimation. Similarly, it is not recommended to use any calibrators having poor antenna field rotation angle coverage for many stations or low S/Ns for both the initial D-term estimation and instrumental polarization self-calibration procedures. They would naturally degrade the D-term estimates, affecting the estimates in the next steps and resulting in poor final D-term estimates.

**Appendix B**

**Validation of Using Calibrators Having Complex Linear Polarization Structures**

In this appendix, we address the question of whether including calibrators having complex linear polarization structures for instrumental polarization self-calibration can improve or degrade the D-term estimates. We consider a realistic case that there is a calibrator that satisfies the similarity assumption reasonably well, and there are two other calibrators having complex linear polarization structures. We produce simulated data with 3C 273–like uv coverage by using PolSimulate, as explained in Section 4.1. We assume that the source’s total intensity emission consists of five point sources, and four of them are linearly polarized (left panel of Figure 12). The Stokes $Q$ and $U$ emission also consists of several point sources, and their positions are the same as the total intensity model components’ positions, except for the one at the origin of the map. For that component, there is a small shift in the positions between the total intensity and linear polarization models by $\approx 0.071$ mas, which is less than one-tenth of the synthesized beam size. This kind of calibrator would normally be treated as a good calibrator that satisfies the similarity assumption well.

We ran GPCAL considering two cases. One is to use these data for the initial D-term estimation using the similarity assumption and perform an additional 10 iterations of instrumental polarization self-calibration using the same data. The other is to repeat the same procedure but including the simulated data with OJ 287 and BL Lac–like uv coverages having significant shifts between the Stokes $I$ and linear polarization model components, which were used for the test of the $P \gg 1$ case in Section 4.1. We compare the reconstructed D-terms, obtained by (i) the initial D-term estimation using the 3C 273 data, (ii) an additional 10 iterations of instrumental polarization self-calibration using the 3C 273 data, and (iii) an additional 10 iterations of instrumental polarization self-calibration using all three data sets simultaneously, with the ground-truth D-terms in Figure 12.

The initial D-term estimates have an RMSE of $\approx 0.5\%$. This result demonstrates that even a very small positional shift ($\approx 0.071$ mas) between the total intensity and linear polarization peaks can prevent an accurate D-term estimation. The RMSE values become smaller with the additional instrumental polarization self-calibration, reaching $\approx 0.12\%$ after 10 iterations when only the 3C 273 data are used. A smaller RMSE value of $\approx 0.093\%$ is obtained when we include two more sources having the linear polarization structures that are significantly different from the total intensity structures for instrumental polarization self-calibration. This test suggests that it is generally recommended to use as many calibrators as possible even though the calibrators have complex linear polarization structures. However, as we suggest in Appendix A, this statement may be valid only when one uses calibrators that satisfy the similarity assumption reasonably well for the initial D-term estimation. Also, one should keep in mind that this result is based on the simulated data assuming simple source structures and no antenna gain errors. As pointed out in Section 4.2.1, if calibrators having poor antenna field rotation angle coverage, low S/Ns, and antenna gain errors that are not well corrected are used, then they will degrade the D-term estimates. It is recommended that users try different combinations of calibrators for the initial D-term estimation and instrumental polarization self-calibration, check the results, and determine the lists of the best calibrators.

**Figure 12.** Color map (left panel): same as Figure 2 but for the simulated data with 3C 273–like uv coverage, produced as explained in Appendix B. Both the total intensity and linear polarization emission consist of several point sources. The total intensity and linear polarization models are slightly shifted from each other by $\approx 0.071$ mas for the component at the origin of the map (“core”), while the models are coincident for the rest of the components (“knots”). Diagrams (right panels): comparison between the ground-truth D-terms assumed in the simulated data and the reconstructed D-terms by GPCAL. The results for three cases are shown: the reconstructed D-terms obtained by the initial D-term estimation using the 3C 273 data (left), an additional 10 iterations of instrumental polarization self-calibration using the 3C 273 data (middle), and an additional 10 iterations of instrumental polarization self-calibration using all three data sets (right). The RMSE values are noted at the top left of each figure. The black dashed lines show one-to-one lines.
References

Asada, K., Inoue, M., Uchida, Y., et al. 2002, PASJ, 54, L39
Astropy Collaboration, Price-Whelan, A. M., Sipőcz, B. M., et al. 2018, AJ, 156, 123
Astropy Collaboration, Robitaille, T. P., Tollerud, E. J., et al. 2013, A&A, 558, A33
Attridge, J. M., Wardle, J. F. C., & Homan, D. C. 2005, ApJL, 633, L85
Casadio, C., Krichbaum, T., Marscher, A., et al. 2017, Galax, 5, 67
Cotton, W. D. 1993, AJ, 106, 1241
Cotton, W. D. 1995a, in ASP Conf. Ser., 82, Very Long Baseline Interferometry and the VLBA, ed. J. A. Zensus, P. J. Diamond, & P. J. Napier (San Francisco, CA: ASP), 189
Cotton, W. D. 1995b, in ASP Conf. Ser., 82, Very Long Baseline Interferometry and the VLBA, ed. J. A. Zensus, P. J. Diamond, & P. J. Napier (San Francisco, CA: ASP), 289
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019a, ApJL, 875, L1
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019b, ApJL, 875, L2
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019c, ApJL, 875, L3
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019d, ApJL, 875, L4
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019e, ApJL, 875, L5
Event Horizon Telescope Collaboration, Akiyama, K., Alberdi, A., et al. 2019f, ApJL, 875, L6
Gabuzda, D. C., Murray, É., & Cronin, P. 2004, MNRAS, 351, L89
Gómez, J. L., Lobanov, A. P., Bruni, G., et al. 2016, ApJ, 817, 96
Gómez, J.-L., Marscher, A. P., Alberdi, A., Jorstad, S. G., & García-Miró, C. 2000, Sci, 289, 2317
Greisen, E. W. 2003, in Information Handling in Astronomy - Historical Vistas (Vol. 285) ed. A. Heck (Dordrecht: Kluwer Academic Publishers), 109
Hada, K. 2017, Galax, 5, 2
Hada, K., Kino, M., Doi, A., et al. 2016, ApJ, 817, 131
Högblom, J. A. 1974, A&As, 15, 417
Homan, D. C., & Lister, M. L. 2006, AJ, 131, 1262
Homan, D. C., & Wardle, J. F. C. 1999, AJ, 118, 1942
Hovatta, T., Lister, M. L., Aller, M. F., et al. 2012, AJ, 144, 105
Hovatta, T., O’Sullivan, S., Martí-Vidal, I., Savolainen, T., & Tchekhovskoy, A. 2019, A&A, 623, A111
Hunter, J. D. 2007, CSE, 9, 90
Issaoun, S., Johnson, M. D., Blackburn, L., et al. 2019, ApJ, 871, 30
Jorstad, S. G., Marscher, A. P., Lister, M. L., et al. 2005, AJ, 130, 1418
Jorstad, S. G., Marscher, A. P., Morozova, D. A., et al. 2017, ApJ, 846, 98
Jorstad, S. G., Marscher, A. P., Stevens, J. A., et al. 2007, AJ, 134, 799
Kettenis, M., van Langevelde, H. J., Reynolds, C., & Cotton, B. 2006, in ASP Conf. Ser., 351, Astronomical Data Analysis Software and Systems XV, ed. C. Gabriel et al. (San Francisco, CA: ASP), 497
Kim, J.-Y., Krichbaum, T. P., Broderick, A. E., et al. 2020, A&A, 640, A69
Kratzchenko, E. V., Kovalev, Y. Y., & Sokolovsky, K. V. 2017, MNRAS, 467, 83
Leppanen, K. J., Zensus, J. A., & Diamond, P. J. 1995, AJ, 110, 2479
Lisakov, M. M., Kovalev, Y. Y., Savolainen, T., Hovatta, T., & Kutkin, A. M. 2017, MNRAS, 468, 4478
Lister, M. L., Aller, M. F., Aller, H. D., et al. 2018, ApJS, 234, 12
Lister, M. L., & Homan, D. C. 2005, AJ, 130, 1389
Marscher, A. 2016, Galax, 4, 37
McKinney, W. 2010, in Proc. 9th Python in Science Conf. 445, Data Structures for Statistical Computing in Python, ed. S. vander Walt & J. Millman (Austin, TX: Scipy), 51
McMullin, J. P., Waters, B., Schiebel, D., Young, W., & Golap, K. 2007, in ASP Conf. Ser., 376, Astronomical Data Analysis Software and Systems XVI, ed. R. A. Shaw, F. Hill, & D. J. Bell (San Francisco, CA: ASP), 127
Nagai, H., Fujita, Y., Nakamura, M., et al. 2017, ApJ, 849, 52
O’Sullivan, S. P., & Gabuzda, D. C. 2009, MNRAS, 393, 429
Park, J., Hada, K., Kino, M., et al. 2019, ApJ, 871, 257
Park, J., Kam, M., Trippe, S., et al. 2018, ApJ, 860, 112
Roberts, D. H., Wardle, J. F. C., & Brown, L. F. 1994, ApJ, 427, 718
Schwab, F. R., & Cotton, W. D. 1983, AJ, 88, 688
Shepherd, M. C. 1997, in ASP Conf. Ser., 125, Astronomical Data Analysis Software and Systems VI, ed. G. Hunt & H. Payne (San Francisco, CA: ASP), 77
Sokoloff, D. D., Bykov, A. A., Shukurov, A., et al. 1998, MNRAS, 299, 189
Tripp, S., Neri, R., Krips, M., et al. 2010, A&A, 515, A40
van der Walt, S., Colbert, C. S., & Varoquaux, G. 2011, CSE, 13, 22
Virtanen, P., Gommers, R., Oliphant, T. E., et al. 2020, Nat. Methods, 17, 261
Wardle, J. F. C., Homan, D. C., Ojha, R., & Roberts, D. H. 1998, Natur, 395, 457
Zavala, R. T., & Taylor, G. B. 2005, ApJL, 626, L73