The Journal of Open Source Software

# An automated pipeline for LOFAR very-long baseline interferometry

**Matthijs van der Wild**[1¶], **Frits Sweijen**[1], **Jurjen M. G. H. J. de Jong**[2,3], **Alexander Drabent**[4], **Emmy L. Escott**[1], **Neal Jackson**[5], **Marcel Loose**[3], **Vijay Mahatma**[6,7], **Leah K. Morabito**[1,8], and **James Petley**[2]

**1** Centre for Extragalactic Astronomy, Deparment of Physics, Durham University, South Road, Durham DH1 3LE, UK **2** Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands **3** ASTRON, Oude Hoogeveensedijk 4, 7991 PD Dwingeloo, The Netherlands **4** Thüringer Landessternwarte, Sternwarte 5, 07778 Tautenburg, Germany **5** University of Manchester, Jodrell Bank Centre for Astrophysics, Department of Physics and Astronomy, Oxford Rd, Manchester M13 9PL, United Kingdom **6** Cavendish Laboratory - Astrophysics Group, University of Cambridge, 19 JJ Thomson Avenue, Cambridge CB3 0HE, United Kingdom **7** Kavli Institute for Cosmology, University of Cambridge, Madingley Road, Cambridge CB3 0HA, United Kingdom **8** Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, United Kingdom ¶ Corresponding author

**DOI:** 10.xxxxxx/draft

**Software**

- Review
- Repository
- Archive

**Editor:**

**Submitted:** 02 December 2025
**Published:** unpublished





## Summary

The Very-Long Baseline Interferometry (VLBI) Pipeline for the International Low-Frequency ARray Telescope (PILoT) is an automated data reduction pipeline that produces calibrated radio data suitable for sub-second resolution imaging. It is a tool that facilitates calibrator and source selection, self-calibration of data, and both postage-stamp and widefield imaging.

While a diverse ecosystem of processing and imaging tools exists for the LOFAR telescope, none of those tools have been designed with high-resolution imaging in mind. As a result, data reduction with the International LOFAR Telescope is a manual and error-prone process. Furthermore, owing to the distributed nature of software development in the LOFAR community, all of these tools have been developed with different input and output conventions.

PILoT aims to incorporate these diverse software tools into a unified framework, making VLBI imaging with LOFAR accessible to a larger group of astronomers. Special care has been placed on ensuring that PILoT is interoperable and reusable, and that all of its software components are controlled through a consistent framework and that intermediate steps of the pipeline can be consistently and safely resumed in the event of intermediate failure. Since typical observations done with the International LOFAR Telescope result in datasets consisting of terabytes of data, the pipeline has been designed to integrate with job schedulers common in High-Performance Computing (HPC) clusters. This minimises manual intervention and optimises the use of available computing resources.

## Statement of need

The International Low-Frequency ARray Telescope (ILT) (van Haarlem et al., 2013) currently comprises 38 Dutch stations and 14 international stations located in partner countries across Europe. It is a radio telescope operating at radio frequencies under 240 MHz with a sensitivity of up to 3 orders of magnitude better than previous telescopes operating at comparable frequencies. By combining data from all stations with baselines of up to 1980 km, the ILT is effectively a continent-sized telescope which is able to image astronomical radio sources at sub-arcsecond resolution (Leah K. Morabito et al., 2025; Varenius et al., 2015).

The VLBI Pipeline for the International LOFAR Telescope (PILoT) is an implementation of a data reduction pipeline which was designed to exploit the full imaging power of the ILT (L. K. Morabito et al., 2022). PILoT addresses several critical issues the original reference implementation had to face:

- The original pipeline was implemented in an obsolete framework, which makes it difficult to impossible to ensure that it would be functional on modern computing infrastructure. In contrast, PILoT is implemented in the Common Workflow Language (CWL; Crusoe et al. (2022)), which is an actively maintained framework in widespread use. This ensures that the pipeline will be logically consistent and maintainable in the long term.
- The original pipeline did not support modern scheduling systems such as Slurm or TORQUE. The implementation in CWL allowed for optimisation of PILoT for the workflow runner toil (Vivian et al., 2017), providing native support for these schedulers. This reduces the runtimes of the pipeline by orders of magnitudes as individual processing jobs can automatically be distributed to available nodes.

In addition, PILoT includes expanded functionality featuring implementations of state-of-the-art advances in imaging techniques such as improvements in imaging resolution (F. Sweijen et al., 2022; Ye et al., 2024), source selection (de Jong et al., 2024; F. Sweijen et al., 2022), and wide-field imaging techniques (de Jong et al., 2025; F. Sweijen et al., 2022).

PILoT forms a natural part of the LOFAR software landscape and is designed to be used on data that has been corrected for various instrumental and ionospheric effects (de Gasperin et al., 2019) and calibrated for directional effects using the data obtained by the Dutch stations using pipelines such as DDF-pipeline (Hardcastle et al., 2019; Tasse et al., 2021) or Rapthor (Rafferty et al., 2021). It uses DP3 (Dijkema et al., 2023), WSclean (Offringa et al., 2014), AOflagger (Offringa, 2010), and the LOFAR Initial Calibration (LINC) pipeline (Drabent et al., 2022), which are developed by the ILT host institute ASTRON, as well as various codebases developed by researchers in the LOFAR community such as the DDF-pipeline (Tasse et al., 2021) for direction-dependent calibration, LOFAR facet-selfcal (R. van Weeren, 2022; R. J. van Weeren et al., 2021) for self-calibration, and the LOFAR Helpers (de Jong, 2021; de Jong et al., 2022) auxiliary library. Finally, it has been adapted and integrated into the FLoCs LOFAR containers (Frits Sweijen et al., 2025) to ensure portability across computing facilities.

PILoT embeds the software tools above into a single cohesive framework and provides several complementary modes with the aim of being a modular and fully automated imaging pipeline. It is able to determine when intermediate results are no longer needed, and disposes of intermediate results once they are no longer required. Furthermore, since the pipeline is optimised to work with toil, processing steps can easily be resumed should the execution of a mode be interrupted for any reason.

The pipeline provides the following main modes of operation:

## Postage stamp imaging

The pipeline supports single-source imaging, in which the calibration is performed on an in-field calibrator to correct for direction-independent phases and delays from the international stations. PILoT selects the best calibrator based on the images and solutions from all performed self-calibration cycles. Following this, multiple imaging targets can be specified at once; the pipeline performs self-calibration and imaging on each target (with a resolution of 0.3 arcseconds) in parallel. The data products are the calibrated data and images and the phase solutions in a H5parm format[1].

[1]The H5parm format is described in appendix C of (de Gasperin et al., 2019).

## Wide-field imaging

The pipeline is capable of intermediate (1–2 arcseconds) and high-resolution (sub-arcsecond) imaging. Before imaging the pipeline removes radio sources from observation data beyond a square field of view of 6.25 degrees$^2$ centred on the imaging target, if given a 6 arcsecond image generated by the DDF-pipeline. High-resolution imaging supports a resolution of 0.6 or 0.3 arcseconds, of which the former reduces the imaging time by a factor of 4 compared to the latter. Intermediate resolution imaging speeds up the imaging time by a factor of 16 compared to the 0.3 arcsecond imaging. The data products are FITS formatted images of the stated resolution.

A list of ongoing research projects where PILoT is a central tool is provided in section 4 of reference (Leah K. Morabito et al., 2025).

# Acknowledgements

MvdW is supported by the Science and Technology Facilities Council via LOFAR-U.K. [ST/V002406/1] and UKSRC [ST/T000244/1]. FS appreciates the support of STFC [ST/Y004159/1]. JMGHJdJ acknowledges support from project CORTEX (NWA.1160.18.316) of research programme NWA-ORC, which is (partly) financed by the Dutch Research Council (NWO), and support from the OSCARS project, which has received funding from the European Commission's Horizon Europe Research and Innovation programme under grant agreement No. 101129751. LKM is grateful for support from a UKRI FLF [MR/Y020405/1] and LOFAR-UK via STFC [ST/V002406/1]. AD acknowledges support by the BMBF Verbundforschung under the grant 05A23STA. ELE is grateful for support from the Medical Research Council [MR/T042842/1]. This research made use of the University of Hertfordshire high-performance computing facility and the LOFAR-UK computing facility located at the University of Hertfordshire and supported by STFC [ST/P000096/1].

The scripts developed for PILoT make use of the ASTROPY (Astropy Collaboration et al., 2022), casacore (Casacore Team, 2019), LoSoTo (de Gasperin et al., 2024), numpy (Harris et al., 2020), pandas (McKinney, 2010), and PyBDSF (Mohan & Rafferty, 2015) libraries.

# References

Astropy Collaboration, Price-Whelan, A. M., Lim, P. L., Earl, N., Starkman, N., Bradley, L., Shupe, D. L., Patil, A. A., Corrales, L., Brasseur, C. E., Nöthe, M., Donath, A., Tollerud, E., Morris, B. M., Ginsburg, A., Vaher, E., Weaver, B. A., Tocknell, J., Jamieson, W., … Astropy Project Contributors. (2022). The Astropy Project: Sustaining and Growing a Community-oriented Open-source Project and the Latest Major Release (v5.0) of the Core Package. *935*(2), 167. https://doi.org/10.3847/1538-4357/ac7c74

Casacore Team. (2019). *casacore: Suite of C++ libraries for radio astronomy data processing*. Astrophysics Source Code Library, record ascl:1912.002.

Crusoe, M. R., Abeln, S., Iosup, A., Amstutz, P., Chilton, J., Tijanić, N., Ménager, H., Soiland-Reyes, S., Gavrilović, B., Goble, C., & Community, T. C. (2022). Methods included: Standardizing computational reuse and portability with the common workflow language. In *Communications of the ACM* (Vol. 65, pp. 54–63). https://doi.org/https://doi.org/10.1145/3486897

de Gasperin, F., Dijkema, T. J., Drabent, A., Mevius, M., Rafferty, D., van Weeren, R., Brüggen, M., Callingham, J. R., Emig, K. L., Heald, G., Intema, H. T., Morabito, L. K., Offringa, A. R., Oonk, R., Orrù, E., Röttgering, H., Sabater, J., Shimwell, T., Shulevski, A., & Williams, W. (2024). *LoSoTo: LOFAR solutions tool*. Astrophysics Source Code Library, record ascl:2401.006.

de Gasperin, F., Dijkema, T. J., Drabent, A., Mevius, M., Rafferty, D., Weeren, R. van, Brüggen, M., Callingham, J. R., Emig, K. L., Heald, G., Intema, H. T., Morabito, L. K., Offringa, A. R., Oonk, R., Orrù, E., Röttgering, H., Sabater, J., Shimwell, T., Shulevski, A., & Williams, W. (2019). Systematic effects in LOFAR data: A unified calibration strategy. *622*, A5. https://doi.org/10.1051/0004-6361/201833867

de Jong, J. M. G. H. J. (2021). LOFAR helpers. In *GitHub repository*. https://github.com/jurjen93/lofar_helpers; GitHub.

de Jong, J. M. G. H. J., van Weeren, R. J., Botteon, A., Oonk, J. B. R., Brunetti, G., Shimwell, T. W., Cassano, R., Röttgering, H. J. A., & Tasse, C. (2022). Deep study of A399-401: Application of a wide-field facet calibration. *668*, A107. https://doi.org/10.1051/0004-6361/202244346

de Jong, J. M. G. H. J., Veefkind, L., Weeren, R. J. van, Oonk, J. B. R., Schlimbach, R. J., Kampert, D. N. G., Wild, M. van der, Morabito, L. K., Sweijen, F., Offringa, A. R., & Röttgering, H. J. A. (2025). Scalable and robust wide-field facet calibration with LOFAR's longest baselines. https://doi.org/10.1093/mnras/staf1373

de Jong, J. M. G. H. J., Weeren, R. J. van, Sweijen, F., Oonk, J. B. R., Shimwell, T. W., Offringa, A. R., Morabito, L. K., Röttgering, H. J. A., Kondapally, R., Escott, E. L., Best, P. N., Bondi, M., Ye, H., & Petley, J. W. (2024). Into the depths: Unveiling ELAIS-N1 with LOFAR's deepest sub-arcsecond wide-field images. *A&A*, *689*, A80. https://doi.org/10.1051/0004-6361/202450595

Dijkema, T. J., Nijhuis, M., Diepen, G. van, Offringa, A., Krombeen, L., Wever, M. de, Maljaars, J., & Loose, M. (2023). *DP3: Streaming processing pipeline for radio interferometric data*. Astrophysics Source Code Library, record ascl:2305.014.

Drabent, A., Rafferty, D., Mancini, M., Loose, M., Horneffer, A., de Gasperin, F., Iacobelli, M., Orru, E., Adebahr, B., Hardcastle, M., Heald, G., Mandal, S., Roskowinski, C., Sabater Montes, J., Shimwell, T., Sridhar, S., Weeren, R. van, & Williams, W. (2022). LOFAR initial calibration (LINC) pipeline. In *Gitlab repository*. https://git.astron.nl/RD/LINC; ASTRON.

Hardcastle, M., Shimwell, T., Tasse, T., & Williams, W. (2019). DDF pipeline. In *GitHub repository*. https://github.com/mhardcastle/ddf-pipeline; GitHub.

Harris, C. R., Millman, K. J., Walt, S. J. van der, Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., Kerkwijk, M. H. van, Brett, M., Haldane, A., Río, J. F. del, Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

McKinney, Wes. (2010). Data Structures for Statistical Computing in Python. In Stéfan van der Walt & Jarrod Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

Mohan, N., & Rafferty, D. (2015). *PyBDSF: Python Blob Detection and Source Finder*. Astrophysics Source Code Library, record ascl:1502.007.

Morabito, L. K., Jackson, N. J., Mooney, S., Sweijen, F., Badole, S., Kukreti, P., Venkattu, D., Groeneveld, C., Kappes, A., Bonnassieux, E., Drabent, A., Iacobelli, M., Croston, J. H., Best, P. N., Bondi, M., Callingham, J. R., Conway, J. E., Deller, A. T., Hardcastle, M. J., … Zucca, P. (2022). Sub-arcsecond imaging with the international LOFAR telescope - i. Foundational calibration strategy and pipeline. *A&A*, *658*, A1. https://doi.org/10.1051/0004-6361/202140649

Morabito, Leah K., Jackson, N., Jong, J. de, Escott, E., Groeneveld, C., Mahatma, V., Petley, J., Sweijen, F., Timmerman, R., & Weeren, R. J. van. (2025). A decade of sub-arcsecond imaging with the International LOFAR Telescope. *370*(2), 19. https:

//doi.org/10.1007/s10509-025-04406-x

Offringa, A. R. (2010). *AOFlagger: RFI Software*. Astrophysics Source Code Library, record ascl:1010.017.

Offringa, A. R., McKinley, B., Hurley-Walker, N., Briggs, F. H., Wayth, R. B., Kaplan, D. L., Bell, M. E., Feng, L., Neben, A. R., Hughes, J. D., Rhee, J., Murphy, T., Bhat, N. D. R., Bernardi, G., Bowman, J. D., Cappallo, R. J., Corey, B. E., Deshpande, A. A., Emrich, D., … Williams, C. L. (2014). WSClean: an implementation of a fast, generic wide-field imager for radio astronomy. *MNRAS*, *444*(1), 606–619. https://doi.org/10.1093/mnras/stu1368

Rafferty, D., Loose, M., Offringa, A., Dijkstra, A. G., Dijkema, T. J., Wever, M. de, Sweijen, F., & Yatawatta, S. (2021). Rapthor: LOFAR DDE pipeline. In *Gitlab repository*. https://git.astron.nl/RD/Rapthor; ASTRON.

Sweijen, Frits, Kurek, A., Weeren, R. van, & Wild, M. van der. (2025). *tikk3r/flocs: v6.0.0* (Version v6.0.0). Zenodo. https://doi.org/10.5281/zenodo.17038431

Sweijen, F., Weeren, R. J. van, Röttgering, H. J. A., Morabito, L. K., Jackson, N., Offringa, A. R., Tol, S. van der, Veenboer, B., Oonk, J. B. R., Best, P. N., Bondi, M., Shimwell, T. W., Tasse, C., & Thomson, A. P. (2022). Deep sub-arcsecond wide-field imaging of the Lockman Hole field at 144 MHz. *Nature Astronomy*, *6*, 350–356. https://doi.org/10.1038/s41550-021-01573-z

Tasse, C., Shimwell, T., Hardcastle, M. J., O'Sullivan, S. P., Weeren, R. van, Best, P. N., Bester, L., Hugo, B., Smirnov, O., Sabater, J., Calistro-Rivera, G., de Gasperin, F., Morabito, L. K., Röttgering, H., Williams, W. L., Bonato, M., Bondi, M., Botteon, A., Brüggen, M., … Wiaux, Y. (2021). The LOFAR Two-meter Sky Survey: Deep Fields Data Release 1. I. Direction-dependent calibration and imaging. *648*, A1. https://doi.org/10.1051/0004-6361/202038804

van Haarlem, M. P., Wise, M. W., Gunst, A. W., Heald, G., McKean, J. P., Hessels, J. W. T., Bruyn, A. G. de, Nijboer, R., Swinbank, J., Fallows, R., Brentjens, M., Nelles, A., Beck, R., Falcke, H., Fender, R., Hörandel, J., Koopmans, L. V. E., Mann, G., Miley, G., … Zensus, A. (2013). LOFAR: The LOw-Frequency ARray. *Astron. Astrophys.*, *556*, A2. https://doi.org/10.1051/0004-6361/201220873

van Weeren, R. (2022). LOFAR facet-selfcal. In *GitHub repository*. https://github.com/rvweeren/lofar_facet_selfcal; GitHub.

van Weeren, R. J., Shimwell, T. W., Botteon, A., Brunetti, G., Brüggen, M., Boxelaar, J. M., Cassano, R., Di Gennaro, G., Andrade-Santos, F., Bonnassieux, E., Bonafede, A., Cuciti, V., Dallacasa, D., de Gasperin, F., Gastaldello, F., Hardcastle, M. J., Hoeft, M., Kraft, R. P., Mandal, S., … Wilber, A. G. (2021). LOFAR observations of galaxy clusters in HETDEX. Extraction and self-calibration of individual LOFAR targets. *651*, A115. https://doi.org/10.1051/0004-6361/202039826

Varenius, E., Conway, J. E., Martí-Vidal, I., Beswick, R., Deller, A. T., Wucknitz, O., Jackson, N., Adebahr, B., Pérez-Torres, M. A., Chyży, K. T., Carozzi, T. D., Moldón, J., Aalto, S., Beck, R., Best, P., Dettmar, R.-J., van Driel, W., Brunetti, G., Brüggen, M., … White, G. J. (2015). Subarcsecond international LOFAR radio images of the M82 nucleus at 118 MHz and 154 MHz. *574*, A114. https://doi.org/10.1051/0004-6361/201425089

Vivian, J., Rao, A. A., Nothaft, F. A., Ketchum, C., Armstrong, J., Novak, A., Pfeil, J., Narkizian, J., Deran, A. D., Musselman-Brown, A., Schmidt, H., Amstutz, P., Craft, B., Goldman, M., Rosenbloom, K., Cline, M., O'Connor, B., Hanna, M., Birger, C., … Paten, B. (2017). Toil enables reproducible, open source, big biomedical data analyses. *Nature Biotechnology*, 314–316. https://doi.org/https://doi.org/10.1038/nbt.3772

Ye, H., Sweijen, F., Weeren, R. J. van, Williams, W., Jong, J. de, Morabito, L. K., Rottgering, H., Shimwell, T. W., Best, P. N., Bondi, M., Brüggen, M., de Gasperin, F., & Tasse, C.

(2024). 1-arcsecond imaging of the ELAIS-N1 field at 144MHz using the LoTSS survey with the international LOFAR telescope. *691*, A347. https://doi.org/10.1051/0004-6361/202348103