A novel radio imaging method for physical spectral index modelling

Ceccotti, E.; Offringa, A.R.; Koopmans, L.V.E.; Timmerman, R.; Brackenhoff, S.A.; Gehlot, B.K.; ... ; Zaroubi, S.

A novel radio imaging method for physical spectral index modelling

E. Ceccotti [1]* A. R. Offringa,[1,2] L. V. E. Koopmans [1] R. Timmerman,[3] S. A. Brackenhoff [1]
B. K. Gehlot [1] F. G. Mertens [4] S. Munshi,[1] V. N. Pandey,[1,2] R. J. van Weeren,[3] S. Yatawatta [2]
and S. Zaroubi[1,5,6]

[1]*Kapteyn Astronomical Institute, University of Groningen, PO Box 800, 9700 AV Groningen, The Netherlands*
[2]*ASTRON, PO Box 2, 7990 AA Dwingeloo, The Netherlands*
[3]*Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands*
[4]*LERMA, Observatoire de Paris, PSL Research University, CNRS, Sorbonne Université, F-75014 Paris, France*
[5]*Department of Natural Sciences, The Open University of Israel, 1 University Road, PO Box 808, Ra'anana 4353701, Israel*
[6]*Department of Physics, Technion, Haifa 32000, Israel*

ABSTRACT

We present a new method, called 'forced-spectrum fitting', for physically based spectral modelling of radio sources during deconvolution. This improves upon current common deconvolution fitting methods, which often produce inaccurate spectra. Our method uses any pre-existing spectral index map to assign spectral indices to each model component cleaned during the multifrequency deconvolution of WSCLEAN, where the pre-determined spectrum is fitted. The component magnitude is evaluated by performing a modified weighted linear least-squares fit. We test this method on a simulated LOFAR high-band antenna (HBA) observation of the 3C 196 QSO and a real LOFAR HBA observation of the 4C+55.16 FRI galaxy. We compare the results from the forced-spectrum fitting with traditional joined-channel deconvolution using polynomial fitting. Because no prior spectral information was available for 4C+55.16, we demonstrate a method for extracting spectral indices in the observed frequency band using 'clustering'. The models generated by the forced-spectrum fitting are used to improve the calibration of the data sets. The final residuals are comparable to existing multifrequency deconvolution methods, but the output model agrees with the provided spectral index map, embedding correct spectral information. While forced-spectrum fitting does not solve the determination of the spectral information itself, it enables the construction of accurate multifrequency models that can be used for wide-band calibration and subtraction.

Key words: instrumentation: interferometers – methods: data analysis – methods: observational – techniques: image processing – techniques: interferometric – radio continuum: general.

1 INTRODUCTION

Modern radio interferometers reach high spatial resolutions and have large instantaneous bandwidth, allowing the generation of sky images with high dynamic range. To achieve this, a meticulous sky model is needed to calibrate data, correcting for effects that occur along the signal path (e.g. Smirnov 2011). This requires accurate modelling of all bright radio sources within a wide field of view, many of which have complex spatial and spectral structures.

Discrete radio sources, such as supernova remnants and radio galaxies, can be modelled by shapelets, which are basis functions constituted by weighted Hermite polynomials that have an analytical Fourier transform (Refregier 2003; Yatawatta 2011). This property makes such functions attractive, because radio interferometers measure visibilities in the Fourier domain – the so-called *uv*-space – of the image brightness distribution. Therefore, shapelets are becoming increasingly more common in calibration software, such as SAGECAL[1] (Kazemi et al. 2011; Yatawatta 2015, 2016; Yatawatta, Diblen & Spreeuw 2017) and RTS (Mitchell et al. 2008; Riding, Mitchell &

Webster 2017), especially because it is possible to use a single basis set of functions to efficiently model extended emission in the sky (e.g. Gehlot et al. 2022). However, to our knowledge, there is no direct implementation yet of shapelets modelling during the cleaning process of imaging. Such models are often produced afterwards through linear least-squares fitting of the basis functions to the final deconvolved image, using software such as SHAMFI[2] (Line et al. 2020), PYBDSF[3] (Mohan & Rafferty 2015), or SHAPELET_GUI[4] (Yatawatta et al. 2013).

In principle, shapelets can be used in compressive sensing techniques, as pointed out by Dabbech et al. (2015). These emerging techniques are based on convex optimization algorithms and signal sparsity to reconstruct the sky model (e.g. Wiaux et al. 2009; Li, Cornwell & de Hoog 2011; Carrillo, McEwen & Wiaux 2012; Garsden et al. 2015; Birdi, Repetti & Wiaux 2020; Terris et al. 2022), where isotropic undecimated wavelets (Starck, Fadili & Murtagh 2007) are usually used instead of shapelets. Compressive sensing is very promising in capturing the finer source details (e.g. Dabbech

---

*E-mail: ceccotti@astro.rug.nl
[1] https://github.com/nlesc-dirac/sagecal

[2] https://github.com/JLBLine/SHAMFI
[3] https://www.astron.nl/citt/pybdsf
[4] https://github.com/SarodYatawatta/shapeletGUI

et al. 2018, 2022), but at the cost of high computational and memory requirements.

One of the most common alternative methods – and less computationally demanding – is the use of CLEAN-based algorithms (Högbom 1974; Clark 1980; Schwab 1984), specifically multiscale (MS) deconvolution, where a source is modelled as a summation of components with different scales (Cornwell 2008). As shown by Offringa & Smirnov (2017), MS cleaning is particularly efficient when Gaussian components are used, because their Fourier transform is also Gaussian. The combination of Gaussian and point components (i.e. delta-functions) allows one to capture the finer structures of sky sources, but at the cost of a number of components that is usually much higher than using shapelets. Line et al. (2020) show that shapelets models leave lower residuals than MS cleaning on the larger spatial scales when these methods are applied to simulated data of radio galaxies. However, results on real data applications show negligible differences because residuals are dominated by other uncertainties, like calibration errors (e.g. Gehlot et al. 2022). Furthermore, current implementations of shapelets do not allow varying spectral indices across the source, unlike Gaussian and point components, where each component can have a different spectral index value. Some compressive sensing-based methods (e.g. Ferrari et al. 2015; Abdulaziz, Dabbech & Wiaux 2019) and other non CLEAN-based techniques, such as maximum entropy methods (Bajkova & Pushkarev 2011) and Bayesian inference techniques (Junklewitz, Bell & Enßlin 2015), can simultaneously model sky-brightness and spectral features, but they have only been tested on simple cases. Moreover, these do not have the advantage of separating the model into components that can easily be inverted to Fourier space. For this reason, MS deconvolution is the de facto method when accurate modelling is required at different frequencies for wide band observations.

In this context, multifrequency (MF) deconvolution algorithms are important, because they allow fitting polynomial functions during deconvolution to simultaneously infer spectral information and increase the deconvolution accuracy (Sault & Wieringa 1994). Combining MS with MF algorithms allows for accurate cleaning of resolved and diffuse sources, while simultaneously inferring spectral information (Rau & Cornwell 2011; Offringa & Smirnov 2017). However, current fitting methods generate unphysical spectral index values, especially for faint and complex (resolved) sources (Rau, Bhatnagar & Owen 2016). This mainly occurs because real data have systematic errors that can cause the fit to diverge, especially at the band edges, producing inaccurate spectral indices (e.g. Offringa et al. 2016). For this reason, spectral information is usually extracted after the imaging process and can be inserted into calibration models afterwards.

Extracting spectral index maps from in-band observations is challenging. Inaccuracies during calibration, such as flux scale alignment (e.g. Shimwell et al. 2022), or imaging, such as spectral fitting during MF deconvolution (e.g. Heywood et al. 2016), can limit the reliability of in-band spectral indices, with errors often exceeding 20 per cent (e.g. Shimwell et al. 2022). Furthermore, inaccuracy in modelling the primary beam can introduce time-dependent spectral features that are too complex to separate from real source components (e.g. Bhatnagar et al. 2008; Tasse et al. 2013; Cotton & Mauch 2021). For these reasons, even the relatively large bandwidth ratio of 2:1 that is reached by telescopes such as LOFAR[5] (van Haarlem et al. 2013) and the VLA[6] (Perley et al. 2011), is not always enough to properly model spectral features. None the less, a few in-band spectral index analysis have been performed with satisfactory results,

especially for strong sources (e.g. McKean et al. 2016; Arias et al. 2018; Fanaroff et al. 2021; Baghel et al. 2023) and source catalogues, such as GLEAM[7] (Callingham et al. 2017; Hurley-Walker et al. 2017, 2022) and LoTSS[8] (Shimwell et al. 2017, 2019, 2022).

The common way to extract a spectral index map is to use different instruments, where a single frequency-integrated image is considered for each wide band. In this way, any in-band uncertainty is reduced and higher signal-to-noise images can be obtained (e.g. de Gasperin, Intema & Frail 2018; Di Gennaro et al. 2021; Fanaroff et al. 2021; Ignesti et al. 2022; Timmerman et al. 2022a, b). The downside is that the spatial resolution and sensitivity of such different telescopes should be as similar as possible. Applying a baseline cut and/or applying some smoothing kernel is always required to make sure that the same angular scales are sampled by different telescopes, which have a different array layout and operate at different frequencies (e.g. Vollmer et al. 2005). This applies also to the in-band spectral index map. However, in such a case, only the change of the *uv*-coverage with frequency must be taken into account and the strength of this correction is usually lower than the one required for multi-instruments observations. This allows in principle for better capture of the finer structures in the spectral index maps, since the resolution is not excessively downgraded (e.g. Fanaroff et al. 2021).

In this paper, we present a novel method, called forced-spectrum fitting, to transfer spectral information from pre-existing spectral index maps into a sky model directly during the deconvolution. This is performed with the MF algorithm of the WSCLEAN imaging software (Offringa et al. 2014), which cleans the image at all frequencies simultaneously, assuming only a single integrated spatial map at a reference frequency, where the clean components are defined. During this step, the forced-spectrum method assigns a power-law spectral index, based on an initial spectral index map, to each of the cleaned components, rather than clean per frequency channel and derive the spectral index afterwards, since this would be affected by differences in, for example, beam, *uv*-coverage, and flagging. The initial spectral index map could be obtained from a different telescope to start with, or from in-band observations themselves, for example making a weighted average of the sky-brightness within certain regions of the source (a method that we call clustering) to reduce the effect of calibration and deconvolution errors. The data can then be calibrated with the forced-spectrum output model, the spectral index map could be adjusted based on the image residuals, and the process can be repeated, as in self-calibration.

The proposed method promises accurate spectral index modelling without the problems related to common MF fitting, especially for extended sources. In fact, MS–MF deconvolution of these sources is less stable than deconvolving point sources, because the degrees of freedom are higher. Such degrees of freedom are reduced by the forced-spectrum method, increasing the stability of the MS–MF deconvolution. This is particularly important for 21-cm experiments, which aim to detect the redshifted 21-cm line emitted by the neutral hydrogen during the Epoch of Reionization (EoR) and Cosmic Dawn (see e.g. Liu & Shaw 2020, for a review). One of the key aspects of these experiments is the separation of the foreground (Galactic and extragalactic) emission from the redshifted 21-cm line. While the former is expected to be smooth over tens of MHz, the latter rapidly fluctuates over MHz-scales (e.g. Shaver et al. 1999; Jelić et al. 2008). Even with many different techniques that can remove the (residual)

---

[5]Low-Frequency Array, http://www.lofar.org
[6]Very Large Array, https://public.nrao.edu/telescopes/vla

[7]GaLactic and Extragalactic All-sky Murchison Widefield Array, https://heasarc.gsfc.nasa.gov/W3Browse/all/gleamegcat.html
[8]LOFAR Two-metre Sky Survey, https://repository.surfsara.nl/collection/lotss-dr2

foreground emission (e.g. Parsons & Backer 2009; Chapman et al. 2012; Bonaldi & Brown 2015; Mertens, Ghosh & Koopmans 2018; Ewall-Wice et al. 2021), spectrally accurate models are still required to improve the calibration process, where errors must be lower than 0.1 per cent to achieve sufficient dynamic range for measuring the 21-cm power spectrum during the EoR (Mazumder et al. 2022).

In Section 2, we outline the forced-spectrum method and its implementation in the MS and MF deconvolution algorithms of WSCLEAN. In Section 3, we introduce the simulated and real data that we use to test our method, whose results are presented in Section 4. Finally, in Section 5, we discuss the results and draw conclusions.

2 METHOD

In this section, we describe a new deconvolution technique called forced-spectrum fitting, which employs a spectral index map to ensure each clean component has a specific spectral index. By using an input spectral index map with physical values, this method generates models with physical spectral information directly during the deconvolution.[9] This approach overcomes the limitations of conventional fitting methods that can incorporate errors and incomplete data, resulting in clean components with unrealistic spectral indices (see further details Section 2.2).

The forced-spectrum method has been implemented in WSCLEAN as an improvement of its multifrequency (MF) deconvolution. Before describing the forced-spectrum method, we briefly illustrate how the MF deconvolution works, especially in combination with multiscale (MS) deconvolution.

2.1 MS–MF deconvolution in WSCLEAN

WSCLEAN[10] is a fast wide-field imager that uses the w-stacking algorithm to correct the w-term in wide-field radio interferometric imaging (Offringa et al. 2014). Its MS and MF algorithms (Offringa & Smirnov 2017) allow deep imaging, generating images with a high dynamic range from which spectral information can be inferred.

The MS method generates a source model that is a summation of point components and basis functions – such as tapered quadratic or circular Gaussian functions – of different sizes (e.g. Cornwell 2008). This reduces negative artefacts around bright resolved sources and allows for a better recovery of extended sources (Rich et al. 2008). Gaussian functions are often preferred because they have an analytically defined Fourier transform.

The MF deconvolution splits a wide frequency band into output channels (i.e. subsets of narrower bandwidth), which are imaged separately. These images are combined in a frequency-integrated continuum image, which has a higher dynamic range, where the peak-finding is performed during the cleaning process. The brightness of a cleaned component is measured in the output images at each frequency channel at the location of the identified peak, thereby allowing the components to model spectral variations. In WSCLEAN, this approach is called joined-channel deconvolution.

The quality of the integrated image can be improved by using the MF weighting, which grids the weights of all output channels in a single grid to ensure the desired weighting (e.g. uniform) for the integrated image. However, this approach may cause the weights

of individual frequencies to deviate, leading to artificial spectral structures due to potentially large deviations in the synthesized beam over frequency.[11] Therefore, in this study, we do not use MF weighting and instead grid the weights of each output channel on separate grids, which allows us to obtain the desired weighting (e.g. uniform) for each individual image.

In MF deconvolution, a polynomial function can be fitted to the measurements to enforce spectral smoothness. In this case, the brightness subtracted from each individual output image is given by the fitted function. WSCLEAN supports two fitting functions, an ordinary polynomial and a logarithmic polynomial, both requiring as extra parameter the number of terms n (with $n \in \mathbb{Z}^+$) at which the Taylor expansion is truncated, giving a $(n-1)$th-order polynomial. The ordinary polynomial at frequency ν is given by

$$S(\nu) = \sum_{i=0}^{n-1} p_i \left(\frac{\nu}{\nu_0} - 1 \right)^i , \tag{1}$$

where S is the flux density and ν_0 is the reference frequency at which coefficients p_i are evaluated.[12] The -1 within brackets is used to make sure that p_0 gives the flux density value $S(\nu_0)$ at the reference frequency. The logarithmic polynomial of $(n-1)$th-order is given by

$$S(\nu) = \prod_{i=0}^{n-1} 10^{c_i \log_{10}^i(\nu/\nu_0)} , \tag{2}$$

where c_i are the coefficients. Note that, $c_0 = \log_{10} S(\nu_0)$ and $c_1 = \alpha$, where α is the spectral index. When $n = 2$, equation (2) can be rewritten as a power law

$$S(\nu) = S(\nu_0) \left(\frac{\nu}{\nu_0} \right)^\alpha . \tag{3}$$

Whatever function is used, the fitting parameters are used to estimate the flux density at the central frequency ν of each output channel and the smooth components are added to the model. During deconvolution, ordinary polynomial functions are generally preferred over logarithmic polynomials, where negative artefacts, which could be picked up by the cleaning, lead to high values at the edge channels, causing instability.

When the fitting is turned on during the MF deconvolution, WSCLEAN can generate a model catalogue of all the clean components. The catalogue consists of information such as sky coordinates, flux density, spectral shapes, etc. for every clean component (see Appendix A for an example). This means that such a model is not limited by the image pixel scale, thereby extended components do not need to be pixelized and very high resolution models can be obtained with fewer components and hence a smaller data volume. The model catalogue can then be directly used for calibration in software such as DP3[13] (Default Pre-Processing Pipeline; van Diepen, Dijkema & Offringa 2018) or can be rendered to model images with the preferred size and resolution.

2.2 Forced-spectrum fitting

Whereas the spectral fitting in the MF–MS deconvolution is very useful to infer spectral information, it often generates clean components with unphysical and inaccurate spectral indices. This can happen

---

[9]Throughout this paper, we will use the term 'physical' to refer to realistic, not extreme spectral index values that do not defy plasma physics as we currently know.

[10]https://gitlab.com/aroffringa/wsclean

[11]https://wsclean.readthedocs.io/en/latest/mf_weighting.html

[12]Here we define $0^0 = 1$, as WSCLEAN and PYTHON do.

[13]https://dp3.readthedocs.io

either because the frequency bandwidth is too small or because spectral smoothness is forced on data that are non-smooth by some systematic, such as beam sidelobes. Fitting errors propagate through the cleaning process, are absorbed into adjacent components, and finally lead to incorrect spectral indices. Therefore, analysing the spectral information produced in this manner is not recommended.

Forced-spectrum fitting solves this problem. It allows generating models with accurate and physically motivated spectral information directly during the deconvolution. This method uses a pre-existing spectral index map to force spectral indices of each clean component during the deconvolution, fitting a first-order ($n = 2$) logarithmic polynomial function described by equation (2). Since c_1 (and optionally higher terms) are fully constrained by the input spectral index map, only c_0 must be fitted. This regulates the fitting and makes it more stable and faster, reducing the degrees of freedom.

To fit for the most probable value of c_0, a (weighted) linear least-squares fit should be performed. When we define a helper function $f(\nu) = (\nu/\nu_0)^{c_1}$ and $S_0 = S(\nu_0) = 10^{c_0}$, the most probable value $\hat{S}_0$ is given by

$$\hat{S}_0 = \frac{\sum_i w_i s_i f(\nu_i)}{\sum_i w_i f^2(\nu_i)} . \tag{4}$$

Here, the sum is over the frequency direction of one pixel, with frequencies ν_i and weight $w_i = 1/\sigma_i^2$, the inverse variance of measurement s_i. However, the use of this function inside joined-channel deconvolution leads to instability. The reason for this is that, during peak-finding, the peak is selected that maximizes $|\sum_i w_i s_i|$. Because equation (4) is not weighted by $f(\nu_i)$, it can happen that a peak is selected for which the weighted squared difference $w_i(s_i - c_0 f(\nu_i))^2$ is already minimized, causing the fitted value $\hat{S}_0$ to be (almost) zero. As a consequence, the residual image remains (nearly) unchanged after the iteration and the same peak is found in the next iteration. This causes an infinite loop, ultimately resulting in the deconvolution process becoming stuck. We observe that this situation is triggered in almost any reasonably deep deconvolution run (several thousand iterations) when implementing forced-spectrum fitting as in equation (4). Therefore, this is a problem that is faced in any application in practice.

One solution to this problem is to simply not weigh the fitting function by $f(\nu_i)$, and instead fit $\hat{S}_0$ using

$$\hat{S}_0 = \frac{\sum_i w_i s_i}{\sum_i w_i f(\nu_i)} . \tag{5}$$

Effectively, this performs the fit using a modified weight $w_i' = w_i f(\nu_i)$. Because peak-finding now selects the pixel that maximizes the absolute numerator in this equation, each deconvolution iteration selects the pixel that maximizes $|\hat{S}_0|$, thereby guaranteeing that progress is made. The downside of this choice is that the fitted spectrum is not optimally weighted. For example, when deconvolving a steep-spectrum source ($c_1 < 0$), high frequencies will be over-weighted, causing its error (caused by noise, sidelobes, and calibration artefacts) to have a larger effect on $\hat{S}_0$. This effect would be strongest when using a large bandwidth with steep or strongly inverted sources. The effect that sidelobes from other sources have is mitigated by the iterative nature of CLEAN-based deconvolution, which involves 'revisiting' a pixel in later iterations. This does of course not hold for noise and calibration artefacts. As a consequence, the use of forced-spectrum fitting requires a higher sensitivity and calibration quality compared to performing the same fit on an image cube outside of deconvolution. This is of course not so surprising, given that spectral deconvolution has to solve a more complex

problem. We will analyse the accuracy of deconvolution in Section 4.

To also force higher-order terms, such as spectral curvature, function f in equation (5) can be changed to have the form of equation (2) divided by 10^{c_0}, i.e.

$$f(\nu_i) = \prod_{i=1}^{n-1} 10^{c_i \log_{10}^i (\nu/\nu_0)} , \tag{6}$$

where the product is over the number of terms, excluding the first term c_0. This is a straightforward extension that will be investigated in a future paper.

Any kind of spectral index map can be used as input of the forced-spectrum method: it can be extracted either from in-band data from the same telescope, or by combining different bands from different instruments. Moreover, smoothing can be applied to increase the robustness of the spectral index estimates. Each map pixel should have a spectral index value. This means that pixels from empty regions of the sky need a spectral index value. Because such regions will not be cleaned (or at a very low level), an arbitrary value can be assigned. For our testing purpose in Section 4, we decide to assign $\alpha = -0.7$, i.e. the typical value for radio sources emitting by synchrotron radiation, to areas where no spectral index value has been extracted, because we still want to assign realistic prior values to any faint source that may not have been previously cleaned. Such assumption is not strict and different observations may require different values. Spectral information are then transferred from the input map to the output models, where each component in the model catalogue will have the spectral index value of the map pixel where its coordinates lay. Such a physical model can be used for self-calibration (see Section 4.2) or even to calibrate different data sets with accurate spectral information.

2.2.1 Behaviour of forced-spectrum for overlapping components

In the forced-spectrum method, the spectral index of a component is assigned based on its central position. When MS deconvolution is used, extended components, such as Gaussians, will have a constant spectrum across their entire shape, determined from their central position. As a consequence, the resulting spectral index of the model may not exactly follow the input spectral index map, and particular artefacts may arise when components overlap. An example case is where one component at position A receives a spectral index α_A and another nearby component at position B receives a different spectral index α_B. If the shapes of the two components overlap, the model at that location becomes a (weighted) combination of the two spectral indices at their centres.

The result of such overlap is shown in Fig. 1, where we consider a Gaussian, with its peak at position A and spectral index α_A, and a point component at position B with spectral index α_B. This means that the forced-spectrum method assigns α_B to the point component, but the spectral index in B extracted by fitting a power law to the model images at different frequencies would be a mix of α_A and α_B. The total flux density resulting in B is the sum of the contributions of the Gaussian in that pixel, S_G, and of the point component, S_P. As shown in the bottom panel of Fig. 1, the sum of two power laws with different spectral indices is not a power law anymore. In Appendix B, we derive its expected slope α, which is a function of frequency that, when $\alpha_A > \alpha_B$, is bound in the range $\alpha_A \gtrsim \alpha \gtrsim \alpha_B$.

When component overlap occurs, the resulting spectral indices in the pixel-based output model images would not agree with the spectral index map used as input in the forced-spectrum method. If
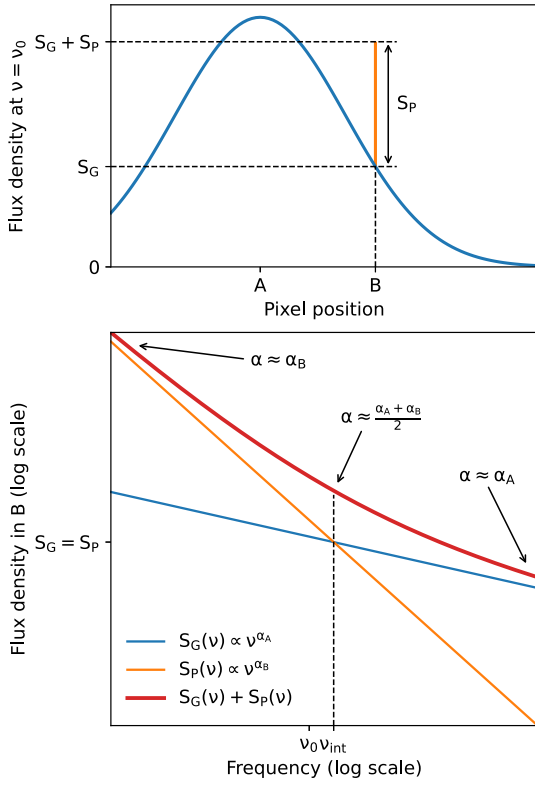
Figure 1. Example case of the overlap of a Gaussian (blue line), with the peak in A and spectral index α_A, and a point component (orange line), located in B with spectral index α_B. The top panel shows such overlap on the image space at a reference frequency ν_0, where the flux densities are S_G and S_P for the Gaussian and the point source, respectively. The bottom panel shows the spectra of the components in the overlapping position B, in a log–log space. The sum of the two components is plotted with the red line, which asymptotically tends to S_P at low frequencies, with a slope $\alpha \approx \alpha_B$, and to S_G at high frequencies, with a slope $\alpha \approx \alpha_A$. Where the single component spectra intersect, i.e. in ν_{int}, the slope of the sum is $\alpha \approx (\alpha_A + \alpha_B)/2$.

the cleaning is performed using only point components, for example, using the Cotton–Schwab algorithm, this issue is not present and the spectral indices resulting from output models exactly match the input map. As we will show in Section 4, the artefacts related to overlapping components are minor compared to the advantages that the combination of forced-spectrum and MS deconvolution provides. In any case, it is important to keep this phenomenon in mind to fully understand our results.

3 DATA SETS

In this section, we discuss the data sets used to test the forced-spectrum method and compare it with traditional polynomial-based fitting. We use a simulated LOFAR high-band antenna (HBA) observation of the 3C 196 quasar (QSO) and a real observation of the FRI radio galaxy 4C+55.16.

LOFAR (van Haarlem et al. 2013) is a low-frequency radio interferometer located in the north of the Netherlands and across Europe. It is constituted by two types of antennas: low-band antennas (LBA), which operate from 10 to 90 MHz, and high-band antennas (HBA), which operate from 110 to 240 MHz. These two antenna types are grouped into stations. Currently, there are 24 core stations (CS; maximum baseline $\sim$ 4 km) near the array centre, 14 remote

Table 1. Observational details of the datasets.

Parameter	Value	
	3C 196[a]	4C+55.16
Telescope	LOFAR HBA	LOFAR HBA
Project code	LC14_020	LC14_019
Antenna configuration	HBA Dual Inner	HBA Dual Inner
Number of stations	75 (NL + IS[b])	28 (ST + RS + IS[b])
Obs. start time (UTC)	2020 Oct 12; 20:30	2020 Nov 9; 23:41
Phase centre (J2000):		
- Right Ascension	$08^h 13^m 36\overset{s}{.}07$	$08^h 34^m 54\overset{s}{.}90$
- Declination	$+48°13'02\overset{''}{.}58$	$+55°34'21\overset{''}{.}07$
Duration of observation	8 h	8 h
Frequency range	110–190 MHz	120–165 MHz
Time resolution[c]	20 s	16 s
Frequency resolution[c]	195 kHz	0.5 MHz

[a]Measurement set into which we predicted 3C 196 simulations.
[b]International station PL611 did not participate in this observation.
[c]After flagging and averaging.

stations (RS; maximum baseline $\sim$ 120 km) across the Netherlands, and 14 international stations (IS; maximum baseline $\sim$ 2000 km) throughout Europe. The international stations allow for very long baseline interferometry (VLBI) observations. Each CS consists of two HBA sub-stations (HBA0 and HBA1) that can be used as independent stations, increasing the number of short baselines and consequently providing a better *uv*-coverage. Additionally, all of the core stations can be combined into a phased-up super-station (ST) that has a much higher sensitivity and a narrower field of view than individual CS. This helps the calibration of international stations and improves the self-calibration, as well as reducing the data volume by about 80 per cent (Morabito et al. 2022).

Both simulations and observations used in this work have been obtained using international stations, reaching sub-arcsecond resolution. The observational details of the two data sets are reported in Table 1. All data processing and analysis is performed on the Dawn high-performance computing cluster (Pandey et al. 2020) at the University of Groningen.

3.1 Simulations of 3C 196

To create complex and realistic simulated data, we take a high-resolution bandwidth-integrated LOFAR model of the 3C 196 QSO (RA $08^h 13^m 36\overset{s}{.}07$, Dec. $+48°13'02\overset{''}{.}58$ in J2000; Spinrad et al. 1985; Pâris et al. 2014) to provide the spatial information, and apply different spectral indices to different parts of the image. The spectral information is taken from a simple, older 4-component model of the source, consisting of Gaussians whose position and shape are reported in Table 2 together with the associated spectral index at a reference frequency of 150 MHz. Despite the morphological simplicity of this model, its spectral indices are consistent with the expected synchrotron emission (Lonsdale & Morison 1980). The source has been divided into two hotspots, northern (N) and southern (S), and two lobes, eastern (E) and western (W), as shown with the dashed contours in Fig. 2. The high-resolution model was obtained from a joined deconvolution of LOFAR–VLBI HBA-mid (116–171 MHz) and HBA-high (230–244 MHz) observations taken in 2016, using MS–MF deconvolution. The model consists of 2812 components (1369 point components and 1444 Gaussians), each one with an unphysical spectral index because of the third-order ordinary polynomial fitting performed during the MF deconvolution. For this reason, we replace the spectral indices of the complex model with

Table 2. Position (RA and Dec. in J2000), spectral index α at 150 MHz, and shape (major and minor axes, and position angle) of the Gaussians of the 4-component model. The position angle (P.A.) is defined relative to the north of the image and not relative to the 'true' north.

Component	RA	Dec	α	Major axis (arcsec)	Minor axis (arcsec)	P.A. (deg)
Hotspot S	$08^h13^m35^s.925$	$48°13'00''.061$	-0.572	1.476	1.091	135.27
Lobe W	$08^h13^m35^s.772$	$48°13'02''.507$	-0.973	2.738	1.754	93.30
Hotspot N	$08^h13^m36^s.182$	$48°13'04''.725$	-0.557	1.034	0.248	124.03
Lobe E	$08^h13^m36^s.389$	$48°13'02''.735$	-0.840	1.456	2.159	324.58

those from the simple model, by linear interpolation of the spectral indices between the centres of the four Gaussian components.

We then render the composite model at two different frequencies, using a pixel scale of 30×30 mas, to extract a pixel-by-pixel spectral index map, using equation (3). During this operation, we only consider pixels brighter than 0.40 mJy beam^{-1}. Since point components generate pixel-scale structures in the spectral index map, we convolve the rendered model images by a circular Gaussian filter with $\sigma = 5$ pixels to obtain a smoother map: the resulting spectral indices span a range between $\alpha_{min} = -0.96$ and $\alpha_{max} = -0.56$. The result is a realistic complex simulated ground-truth model, shown in Fig. 2, which is then used to test our forced-spectrum fitting approach. A gradient between the four regions is visible in the pixel-based spectral index map (right-hand panel), due to overlapping Gaussians, as explained in Section 2.2.1. The ground-truth model of 3C 196 has been generated in the same way that the forced-spectrum method works, namely assigning spectral indices to each component according to its central position.

From the (non-smoothed) ground-truth model, we generate visibilities. We use a realistic *uv*-coverage by predicting the model catalogue into a LOFAR measurement set, phase-rotated to the position of 3C 196. It is an 8 h HBA observation covering 110–190 MHz. All core, remote, and international stations are included to achieve the resolution required to resolve the fine structure of 3C 196. The time and frequency resolutions are 20 s and 195 kHz, respectively. The prediction is performed by DP3. Any flags present in the original measurement set have been removed to ensure the best *uv*-coverage possible.

Finally, we add white noise to each visibility by using the radiometer equation (e.g. Thompson, Moran & Swenson 2017):

$$\sigma_{vis} = \frac{SEFD}{\sqrt{2\Delta t \, \Delta\nu}}, \qquad (7)$$

where SEFD is the system equivalent flux density, $\Delta t = 20$ s is the integration time of a single visibility and $\Delta\nu = 195$ kHz is the frequency channel width. For the sake of simplicity, we assign to each visibility an average SEFD given by a baseline constituted by only core stations, i.e. SEFD = 3.3 kJy (van Haarlem et al. 2013). This value has been increased by 10 per cent to take into account projection effects that can reduce the effective collective area of the antennas when observing away from the zenith. We then find a per-visibility noise level of $\sigma_{vis} = 1.2$ Jy. The results of applying the forced-spectrum method to the simulated data set of 3C 196 will be discussed in Section 4.1.

3.2 4C+55.16 observations

In addition to the simulated observation, we use the FRI radio galaxy 4C+55.16 (RA $08^h34^m54^s.90$, Dec. $+55°34'21''.07$ in J2000; Pilkington & Scott 1965; Charlot et al. 2020), because it has both diffuse and compact emission regions in its lobes and core. We use an 8 h LOFAR–VLBI HBA observation, that have been processed and calibrated by Timmerman et al. (2022b). The data set covers 120–165 MHz, with 16 s and 0.5 MHz of time and frequency resolutions, respectively. All core stations have been phased-up
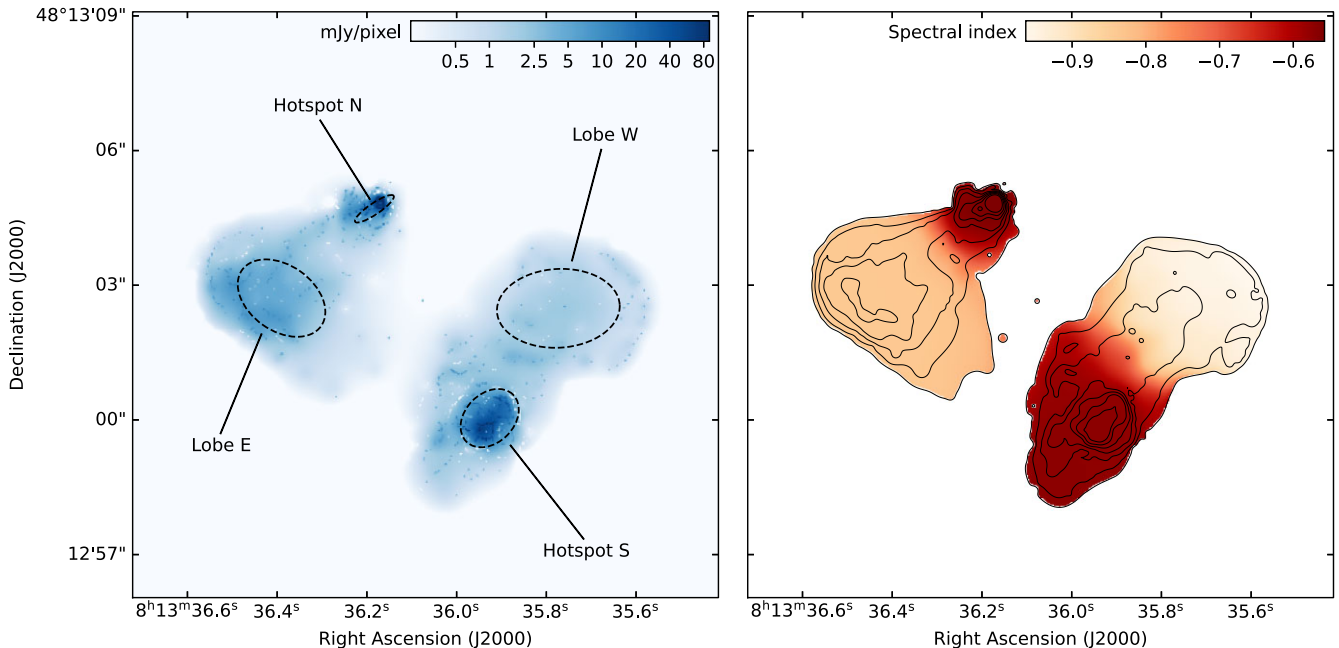

Figure 2. Rendered image of the 3C 196 simulated model at 143 MHz (left) and the ground-truth spectral index map (right) as obtained from the smoothed images. The Gaussians of the simple 4-component model are plotted as dashed ellipses in the left image, with sizes and names as reported in Table 2. The contours in the right figure indicate the model image after the deconvolution with the Gaussian kernel, and they increment in a geometrical progression of 2, starting at 0.40 mJy beam^{-1}. The colour scale of the spectral index map spans in the range $-0.96 \leq \alpha \leq -0.56$.
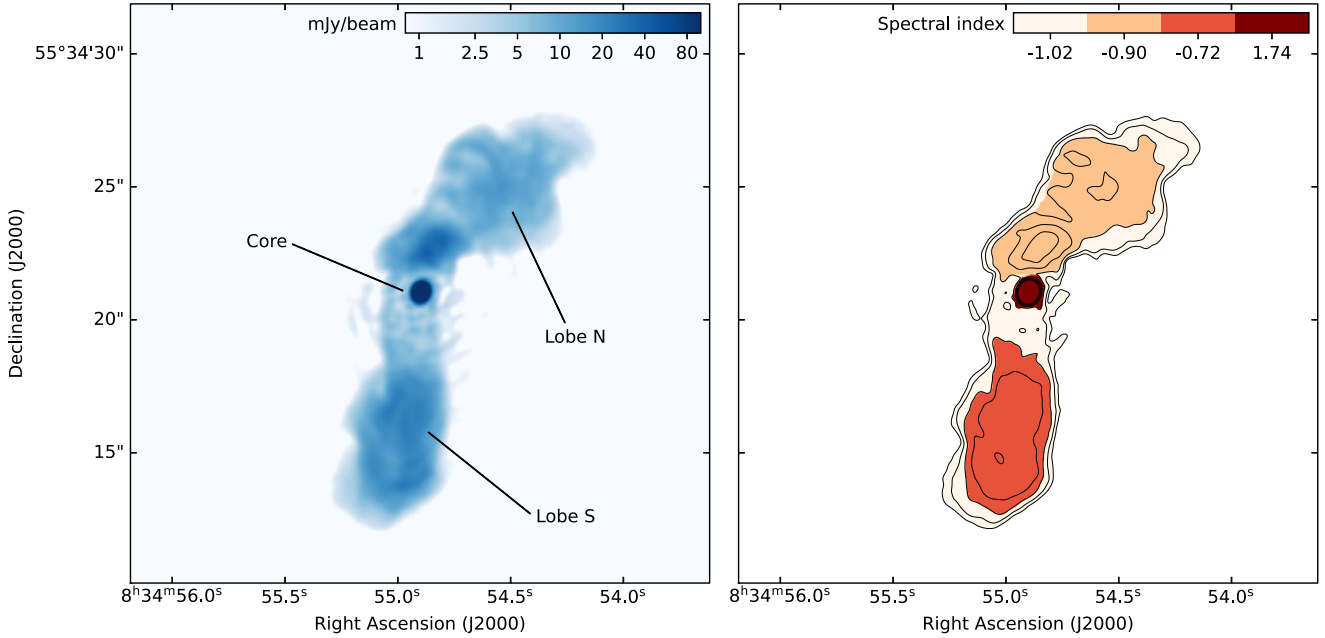
Figure 3. Integrated deconvolved image of 4C+55.16 at 143 MHz (left) and forced-spectrum input spectral index map (right). The image noise is $\sigma = 80 \,\mu\mathrm{Jy} \,\mathrm{beam}^{-1}$. Contours in the right figure are drawn from the integrated image and increment with a factor of 2, starting at 20σ. Spectral index values are extracted from a weighted average of the source brightness within four regions: $\alpha = 1.74$ for the core, $\alpha = -0.72$ for the lobe S, $\alpha = -0.90$ for the lobe N, and $\alpha = -1.02$ for the surrounding low-brightness emission.

into a virtual super-station during the calibration of international stations. This operation reduced the field of view of the array from $\sim 2^\circ$ to ~ 1.5 arcmin, which is sufficient because the source is only ~ 18 arcsec in size.

4C+55.16 has a total flux density of 8.7 Jy at 143 MHz, dominated by the diffuse emission from its lobes that accounts for approximately 60 per cent of the the total emission, while the remaining 40 per cent is concentrated in the core. The core consists of two components with an angular separation of 160 mas, both with inverted spectra ($\alpha > 0$) in the HBA frequency range, but with different spectral indices (Whyborn et al. 1985). This, combined with a steep synchrotron emission from the lobes, produces an integrated spectral index $\alpha = -0.02$ between 143 MHz and 1.4 GHz – as observed by the VLA. The two core components are not completely resolved by LOFAR, but the angular separation is comparable to the angular resolution at the higher frequencies of the observed band. This makes the phase calibration challenging and less accurate, generating imaging artefacts especially when uniform weighting is used. Strong residual sidelobes are observed when MF weighting is not used. To solve this issue, we performed several self-calibration iterations using models from the MS–MF deconvolution of WSCLEAN, fitting a third-order ordinary polynomial function through the output channels. We also normalize the total flux density to the known value of 8.7 Jy at 143 MHz using a power law with $\alpha = -0.02$. Images are made with a pixel scale of 25×25 mas, using uniform weighting without MF weighting, resulting in an integrated synthesized beam of 254.7×183.6 mas when 92 output channels are used. We refer the reader to Appendix C for more details about the self-calibration step.

The next step is to extract a spectral index map that we can use in the forced-spectrum method. We make images to extract the in-band spectral index map for the forced-spectrum method. We use the same WSCLEAN settings as before, except we do not use any kind of fitting because we want the 92 channels to be completely

independent. Using fewer output channels can speed up the fitting process and lower the noise per channel. However, we are currently not limited by computational power and the dynamic range of each image is high enough (>3000). Therefore, we can image each of the 92 narrow channels independently. This minimizes any bias that would occur during the frequency averaging due to non-linear spectral curvature over the averaged channel width. In addition, we apply a circular Gaussian taper with a full width at half-maximum (FWHM) of 250 mas, to make sure that the two core components are not resolved, and we set a circular beam with a size of 300 mas for restoring the clean components at the end of the deconvolution, to have the same resolution over the full bandwidth. Decreasing the resolution with the Gaussian taper increases the image noise, which goes from $67 \,\mu\mathrm{Jy} \,\mathrm{beam}^{-1}$ after the last self-calibration iteration (see Appendix C) to $80 \,\mu\mathrm{Jy} \,\mathrm{beam}^{-1}$ for the obtained integrated image at 143 MHz, shown in the left-hand panel of Fig. 3.

As in Section 3.1, we consider only the brightest emission of the source for the spectral index extraction, removing all the pixels less bright than 20σ. Unfortunately, the *uv*-coverage is reduced when imaging the 92 channels of only 0.5 MHz bandwidth, which causes strong sidelobes. While these sidelobes are not a problem for the self-calibration, because we can set a high cleaning thresholds to not include them in the cleaned model, they strongly affect the pixel-based in-band spectral indices of low surface brightness regions. We thus develop a clustering method to generate in-band spectral index maps. This is based on dividing the source into regions (or clusters) where the emission spectrum is assumed to have the same slope, and assigning the same spectral index to each pixel within a certain region. For 4C+55.16, we divide the source into three regions: the core, the northern (N), and the southern (S) lobes. These have been initially identified by the pixels with brightness higher than 80σ in the integrated image of Fig. 3 and then shaped with small adjustments made by hand. In this case, we have performed these

adjustments to remove a visible sidelobe artefact from the spectral index map area. These changes do not make a significant difference in the results, because the intensity of the image is low at the place of these changes. As a general procedure, it is probably preferred to not make such manual changes, as they may lead to a bias towards what one expects. A fourth region is added to represent the surrounding low brightness emission between 20σ and 80σ.

We then evaluate the weighted mean of the brightness I within each region and per frequency channel, weighting each value by itself, i.e.

$$\overline{I}_{\text{reg}}(\nu) = \frac{\sum_i I_i(\nu)^2}{\sum_i I_i(\nu)} , \qquad (8)$$

where the sum is over the number of pixels of the given region. This kind of average combines more information and reduces the contribution from the fainter components that are most affected by deconvolution artefacts. The spectral indices for each region are finally calculated from the non-linear least-squares fitting of equation (3), where both $S(\nu_0)$ and α are evaluated at $\nu_0 = 150\,\text{MHz}$. In our case, we can directly use the brightness I (units of Jy beam^{-1}) instead of the flux density S (units of Jy) because the beam size is the same for ν and ν_0. The resulting spectral index map is shown in the right-hand panel of Fig. 3, where $\alpha = 1.74$ for the core, $\alpha = -0.72$ for the lobe S, $\alpha = -0.90$ for the lobe N, and $\alpha = -1.02$ for the surrounding low brightness emission. This is the map that we use as input of the forced-spectrum method.

4 RESULTS

In this section, we test the forced-spectrum method on the simulated 3C 196 and the observed 4C+55.16 data sets described in Section 3.

4.1 Results from simulated 3C 196 observation

To be sure that our forced-spectrum method really generates models with physical spectral information, we start testing it with the simulated 3C 196 data set for which we have constructed a ground-truth spectral index map. We make uniform-weighted images of the data with the same pixel scale and size of the ground-truth spectral index map. We take advantage of both MS and MF deconvolution, splitting the full bandwidth into 80 channels, each $\sim 1\,\text{MHz}$, but turning off the MF weighting, similar to Section 3.2. We generate four sets of images with different spectral fitting settings:

(i) first-order ordinary polynomial
(ii) second-order ordinary polynomial
(iii) third-order ordinary polynomial
(iv) forced-spectrum fitting, using the smoothed ground-truth spectral index map as input.

All runs produce an image cube, together with a model and a residual image for each of the 80 channels. We extract spectral index maps from these output models: unlike in Section 3.1, we do not smooth the images, because we want to measure the actual spectral index value that each clean component has – even if it is not fully possible when non-point components are used and hence overlapping occurs, as explained in Section 2.2.1. The extraction is performed as usual, fitting each pixel with the power law of equation (3) at $\nu_0 = 150\,\text{MHz}$, and considering only pixels within the 0.40 mJy beam^{-1} contour level of Fig. 2. While the deconvolved images look similar, the output spectral index maps and residuals are very different, as shown in Fig. 4. It is clear that spectral indices evaluated from ordinary polynomial-fitted models are inaccurate

when compared to the ground-truth spectral index map. Looking for instance at hotspot N, we note that Gaussian components in the centre have much steeper spectra than the initial model. Due to the nature of the CLEAN algorithm, such errors cause higher spectral index values for point components in the same region, resulting in an average value that is ultimately consistent with Table 2 in such a region. In other words, errors during the deconvolution lead to individual components with extreme spectral indices for which the spatially integrated flux matches with the data.

With the forced-spectrum method, the produced spectral index map is much closer to the ground-truth (top-right-hand panel of Fig. 4) – even if they do not exactly match when a pixel-based comparison is performed. This is also evident in Fig. 5, where the pixel distribution of each spectral index map is plotted. The left-hand panel shows the number of pixels with a given spectral index value, binned into bins of 0.01 wide; the right-hand panel shows the pixel-by-pixel difference of ordinary polynomial and forced-spectrum spectral indices, labelled α, with the ground-truth ones, labelled α_{GT}, binned into bins of 0.02 wide. While the left histograms describe the general trend of the output spectral indices with respect to the ground-truth map, the right plot tells us whether α of a given pixel matches α_{GT} of the same pixel, giving an approximate estimate of the fitting error. The ground-truth (grey area in left-hand panel) and the forced-spectrum counts (red line in both panels) have been divided by five to make the comparison with ordinary polynomial-fitted values easier, and all the ordinary polynomial fitting histograms have been averaged into a single distribution in the right-hand panel. The gradient due to overlapping components in the extracted ground-truth map is visible in the left-hand panel as single-sided lobes in the grey distribution. Furthermore, when many components overlap, as for lobe W, peaks of the distribution are shifted with respect to the spectral index values of the four components of 3C 196, marked by the vertical dashed lines. These features occur also for ordinary polynomial and forced-spectrum fitting histograms. Ordinary polynomial fitting of any order generates spectral indices widely spread outside the expected range and with no clearly defined peak. On the other hand, the forced-spectrum distribution follows the ground-truth well.

The minimum and the maximum spectral indices are provided in Table 3 for each fitting method. Considering the full range of the distributions, outlier values increase with the order of the ordinary polynomial. Examining the 95th percentile of the distribution, we notice that forced-spectrum values ($-0.95 \leq \alpha \leq -0.57$) agree with the ground-truth ones ($-0.96 \leq \alpha \leq -0.57$), while ordinary polynomial-fitted spectral indices are outside the expected range. This is confirmed by the right plot of Fig. 5, where the ordinary polynomial fitting histogram reaches values of $|\alpha - \alpha_{\text{GT}}| \approx 0.4$, while, for forced-spectrum, it is confined in the range $|\alpha - \alpha_{\text{GT}}| \lesssim 0.1$, with a sharper peak around zero.

Fig. 4 also shows residual images obtained from the cleaning at the lower, middle, and higher channels, centred at 110.4, 149.4, and 189.5 MHz, respectively. First-order ordinary polynomial fitting generates the largest residuals due to deconvolution errors, with strong positive emission at the edges of the frequency band and negative emission in the middle. This is not surprising because such kind of fitting models the source spectrum with a straight line, while it is a power law by simulation. The residual standard deviation as a function of frequency is shown in the left-hand panel of Fig. 6, where the blue line demonstrate the just described behaviour for the first-order ordinary polynomial function: being high at the edges and in the middle of the band means that the model does not properly reproduce the spectral features of the 3C 196 simulation. On the other hand, the level of the residuals decreases when higher order
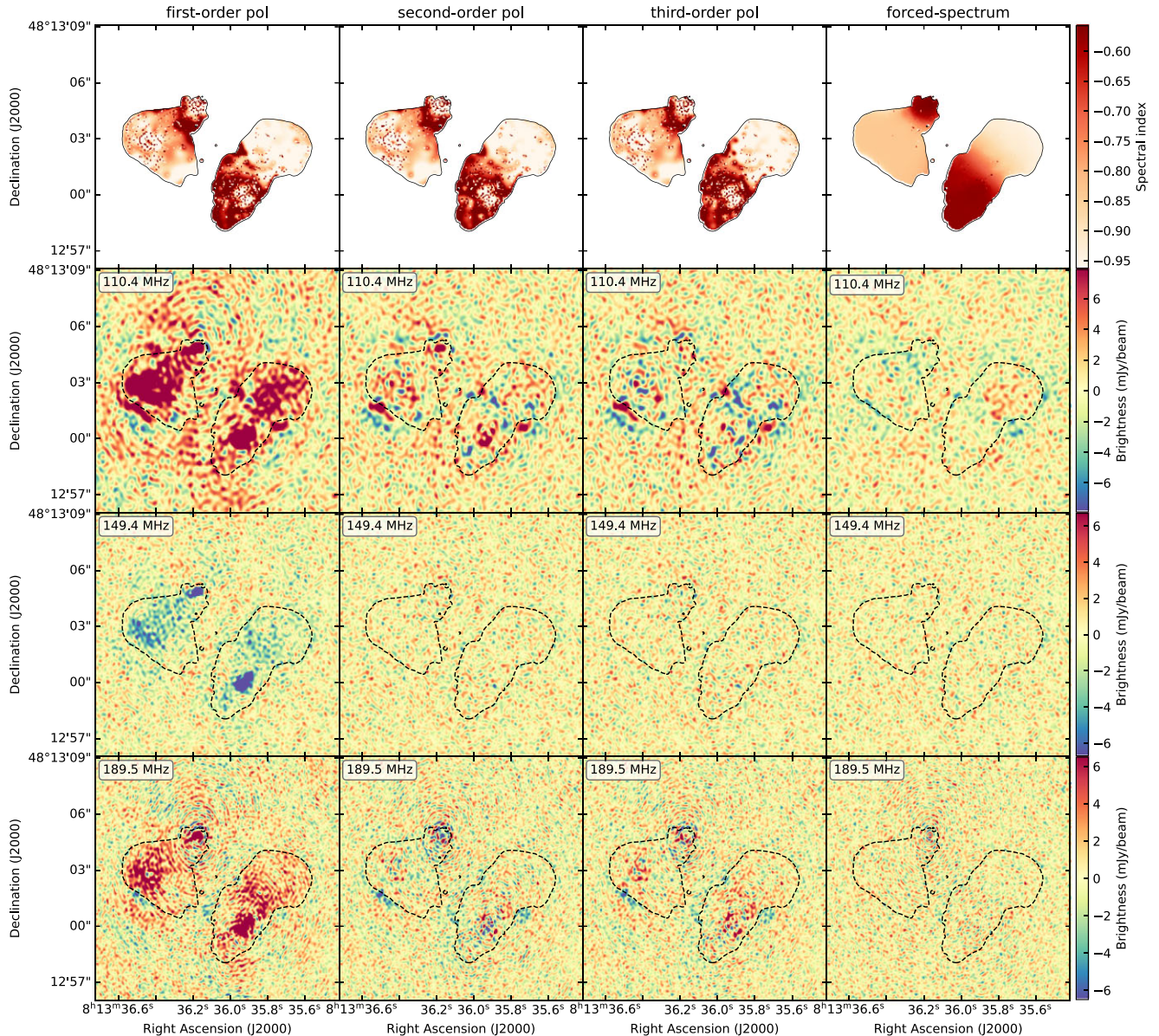
Figure 4. Results from deconvolution with different fitting methods of simulated data of 3C 196. From left to right: results from first, second, and third-order ordinary polynomial fitting, and forced-spectrum fitting. From top to bottom: spectral index maps from output models, and residual images of the lower, middle, and higher channels, centred at 110.4, 149.4, and 189.5 MHz, respectively. The colour scale of spectral index maps is based on the smoothed ground-truth values to make the comparison easier; however, values from ordinary polynomial fits can be outside that range, as shown in Fig. 5 and in Table 3. On the other hand, spectral indices obtained from the forced-spectrum method are similar to the expected values, with only few outlier pixels. The dashed contours superimposed on the residual images indicate the 0.40 mJy beam^{-1} contour level of Fig. 2.

ordinary polynomials are used, because curved lines fit a power law more accurately. The standard deviation is extracted from the region where the source has a brightness higher than 0.40 Jy beam^{-1} (superimposed as dashed contour in Fig. 4). In Fig. 4, especially at 110.4 MHz, we see that second-order ordinary polynomial fitting leaves high residuals in the hotspots and in lobe E, where many point components are present, while a third-order ordinary polynomial more accurately models those components, leaving more uniform residuals. The forced-spectrum method (red line) generates the most uniform and lowest residuals, almost hitting the noise level (black dashed line) at every frequency. Such image noise has been extracted from a Stokes V image and has a frequency-averaged value of $\sigma = 0.65$ mJy beam^{-1}. The standard deviation of the forced-spectrum

residuals is on average 1.3 and 1.2 times lower than second and third-order ordinary polynomials, respectively, and 2.0 times lower than first-order ordinary polynomial (up to 5.7 times lower at 110.4 MHz).

4.1.1 Self-calibration with second-order ordinary polynomial and forced-spectrum models

To test the capabilities of the forced-spectrum output model to improve calibration, we compare the difference between self-calibration with and without the forced spectrum method. The self-calibration is performed with settings that are quite common for LOFAR data sets. In particular, we solve for the diagonal gains of the Jones
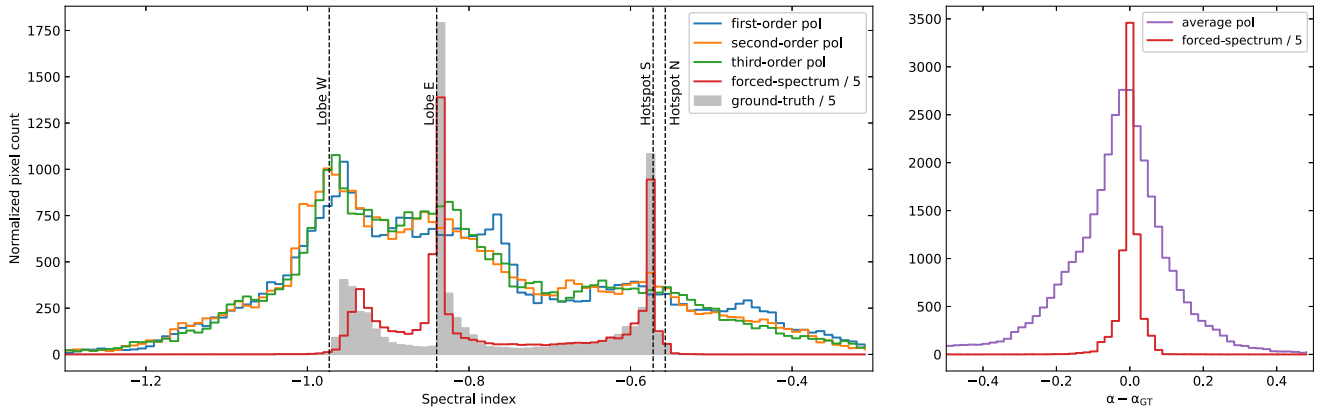
Figure 5. Pixel distributions of the spectral index maps shown in Fig. 4 (left) and of their difference with the non-smoothed ground-truth (α_{GT}) map (right). First (blue), second (orange), and third-order (green) ordinary polynomial fitting distributions are plotted in comparison with the forced-spectrum (red) and the ground-truth (grey area) ones. In the right-hand panel all the polynomial fitting histograms have been averaged into a single distribution (purple). Histograms are obtained by binning spectral indices into bins of 0.01 and 0.02 wide in the left- and right-hand panel, respectively. Forced-spectrum and ground-truth values have been divided by 5 to make the comparison with polynomial distributions easier. Vertical black dashed lines mark the spectral indices of the four components of 3C 196, as reported in Table 2.

Table 3. Minimum (α_{min}) and maximum (α_{max}) values of the spectral index maps obtained from simulated data of 3C 196 for each of the fitting method and for the non-smoothed ground-truth. Both full pixel distribution and 95th percentile ranges are provided. Only pixels within the $0.40\,\text{mJy beam}^{-1}$ contour level of Fig. 2 are considered.

Fitting method	Full range		95th percentile	
	α_{min}	α_{max}	α_{min}	α_{max}
first-order ordinary pol.	-8.94	12.01	-1.28	-0.34
second-order ordinary pol.	-16.05	19.69	-1.36	-0.39
third-order ordinary pol.	-25.30	44.31	-1.37	-0.37
forced-spectrum	-5.47	4.08	-0.95	-0.57
ground-truth	-8.80	0.44	-0.96	-0.57

matrices (i.e. the X and Y gain factors) for each channel and time intervals (i.e. one solution every 195 kHz and 20 s), and calibrate both phase and amplitude. Without the forced spectrum method, the spectral information used for calibration is obtained during the deconvolution, using second-order ordinary polynomial fitting. The calibrated measurement sets are imaged with the same settings as before, and the forced-spectrum fitting is again performed with the smoothed ground-truth spectral index map of Fig. 2.

The results are shown in the right-hand panel of Fig. 6, where the standard deviation of the residuals, labelled σ_{cal}, is subtracted frequency-by-frequency from the uncalibrated one, labelled σ_{uncal}, plotted in the left-hand panel. Whereas $\sigma_{cal} > \sigma_{uncal}$ means that the calibration degrades the image quality, $\sigma_{cal} < \sigma_{uncal}$ does not necessarily mean the opposite, because our already-perfectly simulated data can not be improved by calibration. When this happens, it is likely that the true sky signal or noise is absorbed into the calibration gains.

The second-order ordinary polynomial calibrated data show an overall degradation, especially at low frequencies with first-order ordinary polynomial fitting. Forced-spectrum residuals have $\sigma_{cal} > \sigma_{uncal}$ because we are fitting a power law to data that follow a parabolic curve after the self-calibration step. The line for second-order ordinary polynomials oscillates around $\sigma_{uncal} - \sigma_{cal} = 0$: forcing spectral behaviour that is not intrinsically in the data increases deconvolution errors, leaving such spectral features in the residuals.

This demonstrates the need to calibrate with the best possible model, which must be both spatially and spectrally accurate. Using the forced-spectrum model for the calibration leaves residuals almost unaffected, even if some oscillations are visible because some signal absorption into gains occurs during the calibration.

4.2 Results from 4C+55.16 observation

We test the forced-spectrum method on real observed data to understand how it works when systematics, such as calibration errors, are present. We chose 4C+55.16 as test case, whose self-calibration and spectral index extraction have been described in Section 3.2. We make images following the same procedure used for the 3C 196 simulation in the previous section, but split the bandwidth into 92 channels as we did for the latest self-calibration steps of 4C+55.16. We make a set of images with the first, second, and third-order ordinary polynomial fits, that we compare to the forced-spectrum method using the spectral index map of Fig. 3.

In this section, we will discuss two different results from 4C+55.16: results before and after an extra self-calibration step. To distinguish between the earlier self-calibration steps that were performed, we will refer to this step as the 'final' self-calibration step. The final self-calibration step is performed using the model obtained from the forced-spectrum deconvolution. This will therefore also test whether the use of a forced-spectrum model may improve calibration.

4.2.1 Before final self-calibration

We will start by discussing the results before the final self-calibration. As explained in Section 2.2.1, in the model catalogue (an example from 4C+55.16 is provided in Appendix A), each component has the expected spectral index from the input map based on its coordinates, because components are listed with their central position and spectral index in the model catalogue. However, the residuals from the forced-spectrum method are higher than those from the ordinary polynomial fits. The standard deviation of the 92 residual images are shown in the left-hand panel of Fig. 7 for the four fitting methods, extracted from the 20σ contour level of Fig. 3 after smoothing
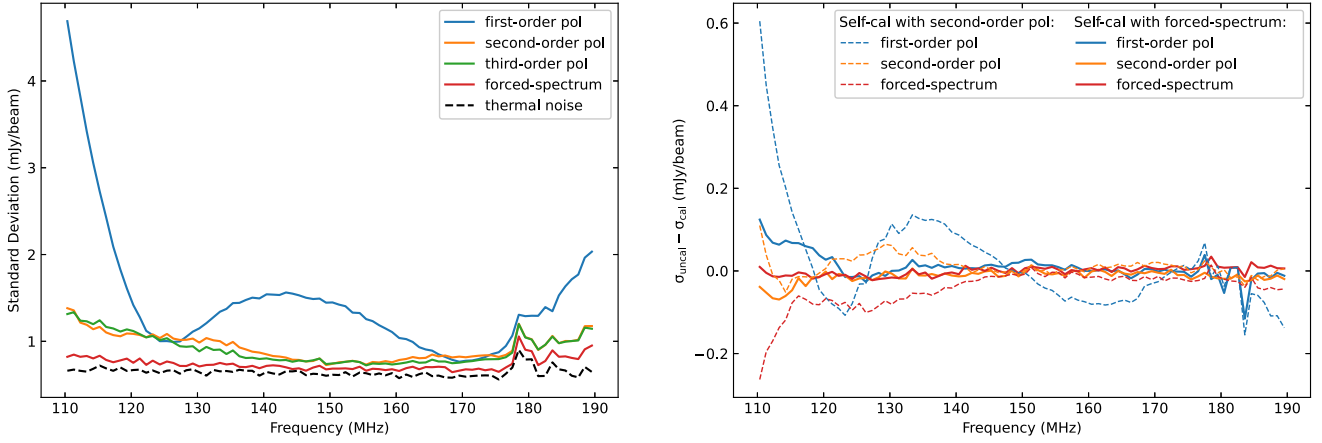
Figure 6. Standard deviation of residuals obtained from deconvolution of initial (σ_{uncal}) simulated 3C 196 data (left) and its difference with the standard deviation after a self-calibration step (σ_{cal}) using different models (right). Different fitting methods have been used: first (blue), second (orange), and third-order (green) ordinary polynomial functions, and forced-spectrum fitting (red). In the right-hand panel, the self-calibration has been performed with the second-order polynomial (dashed lines) and forced-spectrum (solid lines) models obtained from the deconvolution of the initial data. Image thermal noise is drawn in the left-hand panel with a black dashed line. Only residuals within the dashed contours as Fig. 4 are used.
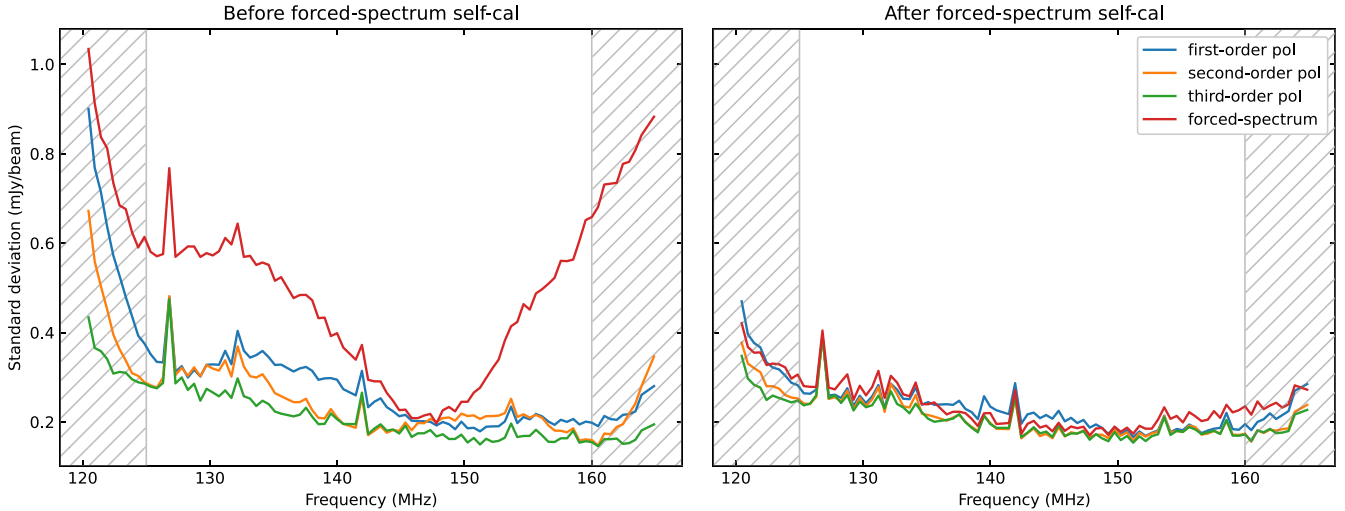

Figure 7. Standard deviation of smoothed residuals obtained from deconvolution of 4C+55.16 data before (left) and after (right) the last self-calibration using the forced-spectrum model. First (blue), second (orange), and third-order (green) ordinary polynomial functions, and forced-spectrum fitting (red) are shown. The hatched areas mark the frequencies excluded from the analysis (we consider the range 125–160 MHz). Only residuals within the 20σ contour level of Fig. 3 are used.

using a circular Gaussian kernel with an FWHM of 500 mas.[14] This reduces the contribution of high resolution errors that arise because the cores are inaccurately modelled; since these are caused by the fundamental resolving resolution of the instrument and not by the spectral modelling, we do not want our residual images to be dominated by those errors. Residuals are higher at the band edges because fitting is less accurate here. For this reason, we will focus our analysis only on the range 125–160 MHz. Third-order ordinary polynomial fitting results in the lowest residuals. This may be caused by the previous self-calibration steps, where we used models obtained by fitting ordinary polynomial functions of such order. However, the other ordinary polynomial results show comparable residuals,

especially at frequencies higher than 146 MHz, in contrast to what we have seen in the simulations (see Fig. 6). First-order ordinary polynomial fitting has the largest residuals, which are on average 1.3 times higher than other ordinary polynomial fitting methods, up to 1.5 times higher at 139.0 MHz. Such small differences highlight how higher-order ordinary polynomial fits do not significantly improve residuals when systematics are present in the data (Offringa et al. 2016).

Unlike in our simulations, the forced-spectrum method produces higher residuals. They are on average 2.1 times higher than third-order ordinary polynomial fitting, with differences getting larger toward both ends of the band, and is 4.2 times higher at 159.5 MHz. Most of such excess residual power comes from the core region, which switches from positive to negative residual brightness from low to high frequency. The (smoothed) maximum residual brightness is 6.6 mJy beam^{-1} at 125.3 MHz, while the minimum

[14] We removed the output channels centred at 136.08, 140.97, and 164.40 MHz because of bad image quality due to interference at these frequencies.

is $-11.1 \, \text{mJy beam}^{-1}$ at 159.5 MHz. Since $\alpha > 0$ for the core, this means that the spectral index estimated for the core is too high. In the range 145–150 MHz, the forced-spectrum is comparable to ordinary polynomial fitting. Furthermore, the excess residual power becomes stronger at higher frequencies. Part of this might be explained by the intrinsic double nature of the core, which we modelled as a single component with a single spectral index value, while we know that the two components have different spectral signatures.

4.2.2 After final self-calibration

We will now discuss the results after the final self-calibration. This step is performed using the forced-spectrum output model, solving for the diagonal gains of the Jones matrices for each channel and time intervals (i.e. one solution every 0.5 MHz and 16 s), and calibrating both phase and amplitude. We make new sets of images, still using the initial input map for the forced-spectrum fitting.

The final results are summarized in Fig. 8. The top row shows the spectral index maps directly extracted (i.e. with no smoothing) from the output models for the four fitting methods, considering only pixels within the 20σ contour level of Fig. 3, this time matching what we have found from the simulations: forced-spectrum method generates more spectrally accurate models than ordinary polynomial fits. Ordinary polynomial-fitted models deviate from the expected values especially at the edges of the source, where the signal to noise is lower. This happens because images in those regions are dominated by calibration and deconvolution artefacts, which can be picked up as clean components during deconvolution. The spectral index of the core is a combination of point components and Gaussians. For some of these Gaussians, their centres fall into the lobes N and S. The extracted spectral index from this underlying diffuse emission is therefore an average between the values of the lobe N and lobe S for the forced-spectrum output models (see Section 2.2.1).

Pixel distributions of the spectral index maps are shown in Fig. 9. The left-hand panel shows the number of pixels with a given spectral index value, whereas the right-hand panel shows the pixel-by-pixel difference of ordinary polynomial and forced-spectrum spectral indices, labelled α, with the input ones, labelled α_{input}, both binned into bins of 0.04 wide. Input map (grey area in the left-hand panel) and forced-spectrum counts (red histograms in both panels) have been divided by 4 in the left-hand panel and by 8 in the right-hand panel, where all the ordinary polynomial fitting histograms have been averaged into a single distribution (purple histograms). Furthermore, the right-hand panel shows two kind of distributions for each fitting method: empty histograms represent pixels with $S(\nu_0) > 80\sigma$ (i.e. core and lobes regions of Fig. 3), while full histograms represent pixels with $20\sigma < S(\nu_0) < 80\sigma$ (i.e. surrounding low brightness emission), where $\sigma = 80 \, \mu\text{Jy beam}^{-1}$. The forced-spectrum method generates spectral indices that agree well with the input map, while ordinary polynomial fits generate values with a larger spread, with $|\alpha - \alpha_{\text{input}}| \approx 0.7$ for the core and lobes regions, reaching differences higher than 1 for the low brightness emission. Ordinary polynomial fits tend to generate $\alpha > \alpha_{\text{input}}$ in low brightness regions, while $\alpha < \alpha_{\text{input}}$ in high brightness ones. This is also visible in Table 4.

Modelling the core with only a few point components causes a smaller peak at the input spectral index value, since most of the core pixels contains the diffuse emission described above. This generates small gradients in the forced-spectrum histogram between the three main peaks. However, such gradients are mainly due to the leaking of the Gaussian components of the lobes into the surrounding low-brightness region. Many pixels from this region take the spectral

index values of lobe N and lobes S, resulting in a number of pixels with those values higher than the input ones. The right-hand panel clearly shows this effect: for fluxes lower than 80σ, two peaks arise at $\alpha - \alpha_{\text{input}} = 0.12$ and 0.30. Since $\alpha_{\text{input}} = -1.02$ in this region, we recover $\alpha = -0.90$ and -0.72, which are the input values for lobe N and lobes S, respectively.

Residual images are also shown in Fig. 8, obtained from a low, middle, and high channels, centred at 125.3, 145.8, and 159.5 MHz, respectively. As before the final self-calibration, we smooth the images with a circular Gaussian filter with an FWHM of 500 mas. Looking at the residuals within the dashed contour line, representing the 20σ level, it is evident that differences between the four fitting methods are negligible in the lobe regions, but not in the core. Residuals from the forced-spectrum fitting still show an excess brightness that is positive at low frequency and negative at high frequency, becoming comparable to second and third-order ordinary polynomial fittings in the middle of the band. However, the self-calibration has reduced the maximum and minimum residual brightness at 125.3 and 159.5 MHz by a factor of ~6, which now are 1.2 and $-1.9 \, \text{Jy beam}^{-1}$, respectively. Extracting the standard deviation as we did for the left-hand panel of Fig. 7, we see that the differences are almost negligible between all the fitting methods, as shown in the right-hand panel of Fig. 7. Forced-spectrum residuals are on average only 1.1 times higher than third-order ordinary polynomial residuals, which are the lowest over most of the band. The strongest difference is toward the high-end of the band, and is of the order of 30 per cent. As expected, first-order ordinary polynomial fitting generates the highest residuals in the middle of the band, because a straight line fails to fit a power law. However, even in this case, the difference is just 24 per cent with the third-order ordinary polynomial. In addition, the overall standard deviation is lower than before the final self-calibration. This means that self-calibrating with the models obtained from the forced-spectrum fitting improves the overall data quality, especially at the low-end of the band.

5 DISCUSSION AND CONCLUSIONS

In this paper, we have presented a novel method – implemented in WSCLEAN – for generating sky models with physical spectral information, by taking advantage of previously created spectral index maps. While the method itself does not generate spectral index information, it allows for the decomposition of convolved data into components. This produces a model that closely matches the data and can be used for further calibration or source subtraction. The input spectral index map is used inside the multifrequency (MF) and multiscale (MS) CLEAN algorithm, and constrains the spectral index term of every found component. As a consequence, only a single scaling factor must be calculated through a modified-weighted linear least-squares fit (Section 2.2). Thus, the forced-spectrum method reduces inaccuracies that are common in typical MF fitting methods, particularly for observations with small bandwidth, and generates accurate spectral models directly during the deconvolution. Any type of spectral index map can be used, whether from multi-instrument observations or in-band data, providing flexibility in how the spectral information that is transferred into the sky model is determined. Because the forced-spectrum method assigns the spectral index to each cleaned component based on its central position, inaccuracies in the pixel-based spectral index maps may result when extended (i.e. Gaussian) components overlap (Section 2.2.1). This spectral index mixing primarily affects the low-brightness parts of a source and generates gradients between regions with different spectral indices. The overlapping issue can be avoided using CLEAN algorithms that
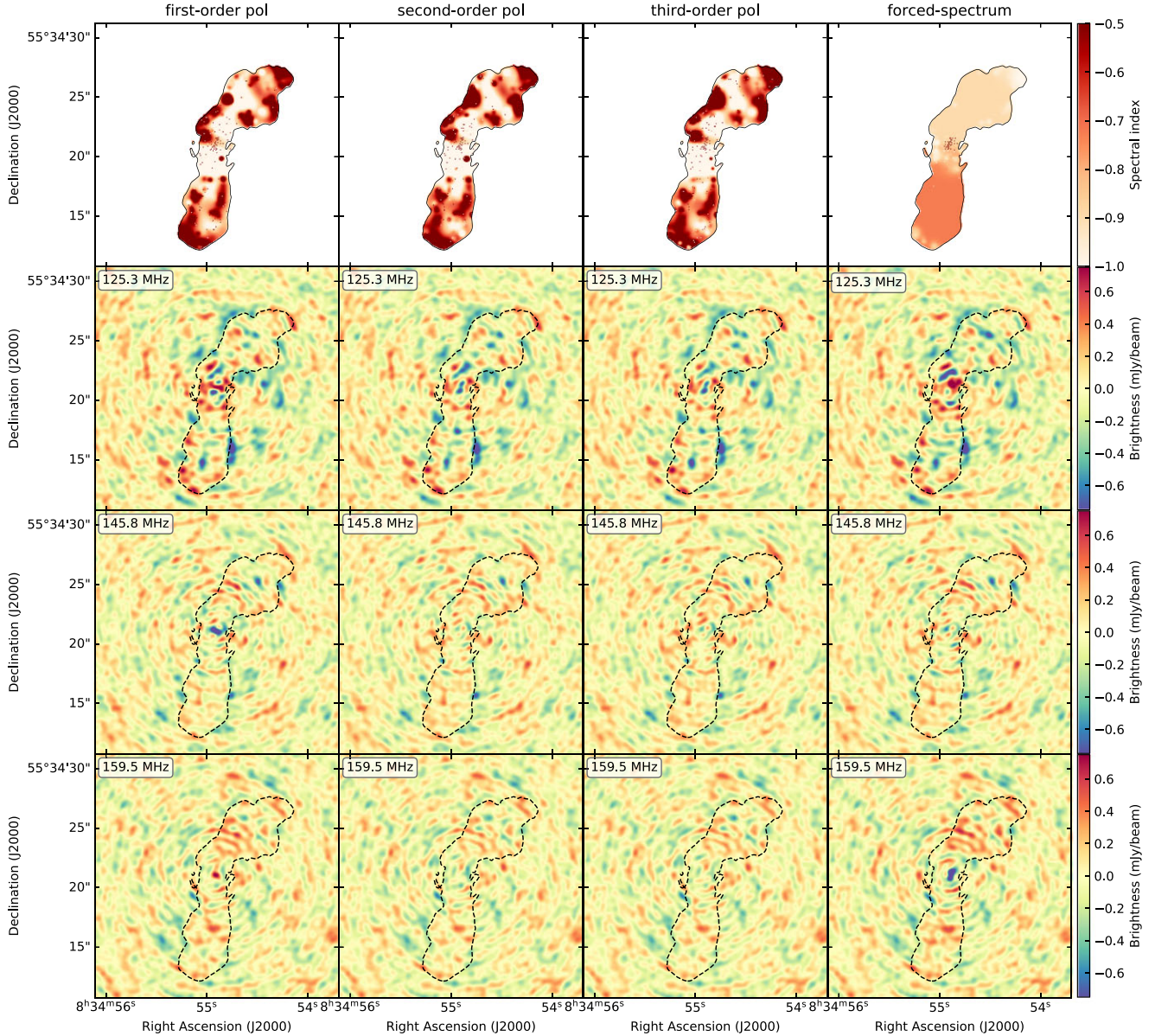
Figure 8. Deconvolution results of 4C+55.16 data with different fitting methods after final self-calibration using forced-spectrum output models. From left to right: results from first, second, and third-order ordinary polynomial fitting, and from the forced-spectrum fitting. From top to bottom: spectral index maps from output models, and residual images of a low, middle, and high channels, centred at 125.3, 145.8, and 159.5 MHz, respectively, avoiding the edges of the observed band. Residuals images have been smoothed by a circular Gaussian kernel with an FWHM of 500 mas to enhance the strong features. The colour scale of spectral index maps is the same of Fig. 3 to make the comparison easier with the forced-spectrum input map; however, values from ordinary polynomial fits can be outside that range, as shown in Fig. 9 and in Table 4. On the other hand, spectral indices obtained from the forced-spectrum method are similar to the expected values with only few outlier pixels. The dashed contours superimposed on the residual images indicate the 20σ level of Fig. 3.

are only based on point components, but the benefits of using the MS deconvolution outweigh this minor issue.

We also demonstrate a clustering method for extracting spectral indices from in-band observations (Section 3.2). The use of MF deconvolution in such observations often generates output channel images that are dominated by calibration and deconvolution errors, making it challenging to determine accurate spectral indices. To overcome this, in the clustering method, we divide the source into a certain number of regions, calculate the weighted average of the brightness within each region, and extract the spectral indices. This reduces the effect of systematics.

Spectrally forced imaging provides a new tool that can be used to improve generic MF imaging. We have focused our testing on the imaging of individual non-varying sources. We have not yet tested wide-field imaging or deconvolution of transients and point sources. The method may improve results for those use-cases as well, but it is likely that its strength lies in the modelling of individual, persistent, and resolved sources that are relatively challenging to model.

The forced-spectrum method works well in combination with the MF–MS deconvolution of WSCLEAN when a model catalogue is also generated. Since such a model is not limited by the pixel scale, it can be used to save high-resolution models of calibrators and strong sources that could be used by different experiments, reducing
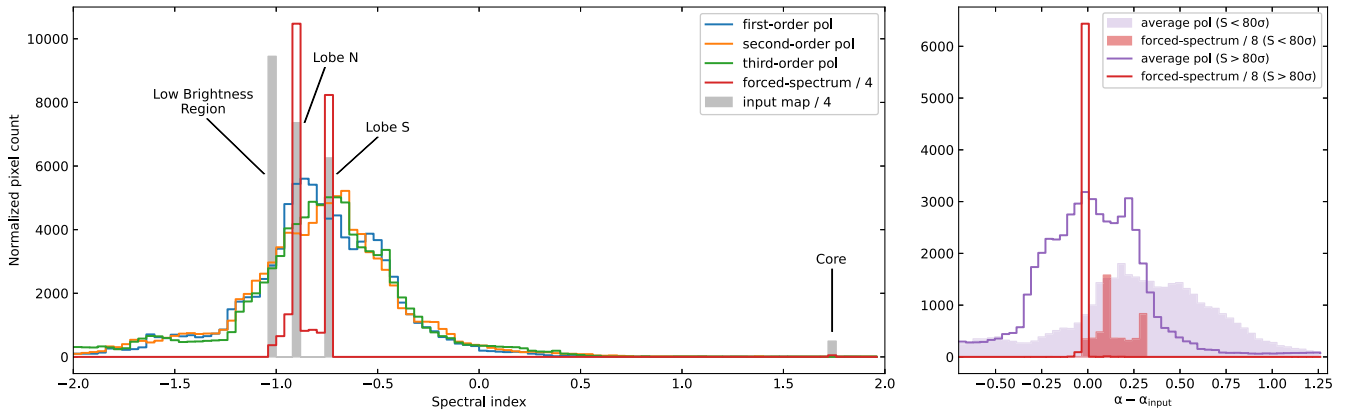
Figure 9. Pixel distributions of the spectral index maps shown in Fig. 8 (left) and of their difference with the input (α_{input}) map (right). First (blue), second (orange), and third-order (green) ordinary polynomial fitting distributions are plotted in comparison with the forced-spectrum (red) and the input map (grey area) ones. In the right-hand panel, all the polynomial fitting histogram have been averaged into a single distribution (purple). Here, empty and filled histograms respectively show distributions of pixels with flux density higher and lower than the 80σ level of Fig. 3. The histograms are obtained by binning spectral indices into bins of 0.04 wide in both panels. Forced-spectrum and input map values have been divided by 4 in the left-hand panel and by 8 in the right-hand panel, to make the comparison with polynomial distributions easier.

Table 4. Minimum (α_{min}) and maximum (α_{max}) values of the spectral index maps obtained from the final self-calibrated data of 4C+55.16 for each of the fitting methods. Both full pixel distribution and 95th percentile ranges are provided. Only pixels within the 20σ contour level of Fig. 3 are considered.

Fitting method	Full range		95th percentile	
	α_{min}	α_{max}	α_{min}	α_{max}
first-order ordinary pol.	-12.27	24.64	-1.82	-0.11
second-order ordinary pol.	-24.26	49.38	-1.70	-0.31
third-order ordinary pol.	-29.48	54.34	-1.96	0.21
forced-spectrum	-1.20	2.92	-0.99	-0.72
input map	-1.02	1.74	-1.02	-0.72

the required data volume. Providing well-motivated external spectral information into such models is crucial for high-resolution and wide-field interferometers operating at low-frequencies, such as LOFAR and the upcoming SKA-Low[15] (Dewdney et al. 2009; Braun et al. 2019), which have the capability to generate images ∼100 000 pixels in size.

Results from both simulations (Section 4.1) and real observations (Section 4.2) have shown an improvement with respect to typical ordinary polynomial fitting methods, which generate inaccurate spectral indices, especially for faint sources. Promising results are also observed in the residual images, which appear more uniform and lower in magnitude in the simulations. On the other hand, real observations are dominated by other systematics, such as calibration and deconvolution errors. These can be overfitted and absorbed by high-order ordinary polynomial fits – into incorrect spectral indices – and eventually produce better residuals than forced-spectrum fitting, which does not allow the same freedom and such systematic spectral features are visible in the residual images. Despite this, we have shown that more accurate models are obtained and that they can be used for calibration, ultimately improving the overall data quality, resulting in residuals that are similar to those obtained with ordinary polynomial fits. The models generated by forced spectral fitting can also be more easily extrapolated to different frequencies. This may prove to be useful for sharing models between instruments.

An interesting other application of the forced-spectrum method could be to determine how well a spectral index map matches the data, by looking at the residuals after forced-spectrum deconvolution. Image residuals may also be used to adjust the input spectral index map, especially if it has been made with the clustering method. For example, if the residual brightness of pixels or regions go from positive (negative) at low frequency to negative (positive) at high frequency, the spectral indices of those components should be steeper (flatter) than the input ones. The adjusted spectral index map better matches the data and can be used to output a model that can then be used to self-calibrate the data set.

These results are highly relevant for improving upper limits on the 21-cm power spectrum from the Epoch of Reionization and Cosmic Dawn. With calibration errors below 0.1 per cent being critical for achieving the high dynamic range needed to measure the 21-cm power spectrum during these cosmic eras (Mazumder et al. 2022), the sky models derived from forced-spectrum fitting hold significant promise for enhancing the accuracy of such measurements.

[15]Square Kilometers Array, https://www.skao.int/en/explore/telescopes

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abdulaziz A., Dabbech A., Wiaux Y., 2019, MNRAS, 489, 1230
Arias M. et al., 2018, A&A, 612, A110
Arras P., Reinecke M., Westermann R., Enßlin T. A., 2021, A&A, 646, A58
Astropy Collaboration et al., 2022, ApJ, 935, 167
Baghel J., Kharb P., Silpa, Ho L. C., Harrison C. M., 2023, MNRAS, 519, 2773
Bajkova A. T., Pushkarev A. B., 2011, MNRAS, 417, 434
Bhatnagar S., Cornwell T. J., Golap K., Uson J. M., 2008, A&A, 487, 419
Birdi J., Repetti A., Wiaux Y., 2020, MNRAS, 492, 3509
Bonaldi A., Brown M. L., 2015, MNRAS, 447, 1973
Braun R., Bonaldi A., Bourke T., Keane E., Wagg J., 2019, preprint (arXiv:1912.12699)
Callingham J. R. et al., 2017, ApJ, 836, 174
Carrillo R. E., McEwen J. D., Wiaux Y., 2012, MNRAS, 426, 1223
Chapman E. et al., 2012, MNRAS, 423, 2518
Charlot P. et al., 2020, A&A, 644, A159
Clark B. G., 1980, A&A, 89, 377
Cornwell T. J., 2008, IEEE J. Sel. Top. Signal Processing, 2, 793
Cotton W. D., Mauch T., 2021, PASP, 133, 104502
Dabbech A., Ferrari C., Mary D., Slezak E., Smirnov O., Kenyon J. S., 2015, A&A, 576, A7
Dabbech A., Onose A., Abdulaziz A., Perley R. A., Smirnov O. M., Wiaux Y., 2018, MNRAS, 476, 2853
Dabbech A., Terris M., Jackson A., Ramatsoku M., Smirnov O. M., Wiaux Y., 2022, ApJ, 939, L4
Dewdney P. E., Hall P. J., Schilizzi R. T., Lazio T. J. L. W., 2009, IEEE Proc., 97, 1482
Di Gennaro G. et al., 2021, A&A, 654, A166
Ewall-Wice A. et al., 2021, MNRAS, 500, 5195
Fanaroff B. et al., 2021, MNRAS, 505, 6003
Ferrari A., Deguignet J., Ferrari C., Mary D., Schutz A., Smirnov O., 2015, preprint (arXiv:1504.06847)
Garsden H. et al., 2015, A&A, 575, A90
Gehlot B. K. et al., 2022, A&A, 662, A97
Gooch R., 1996, in Jacoby G. H., Barnes J., eds, 101, ASP Conf. Ser. Vol.101, Astronomical Data Analysis Software and Systems V. Asron. Soc. Pac., San Francisco, p. 80
Harris C. R. et al., 2020, Nature, 585, 357
Heywood I. et al., 2016, MNRAS, 460, 4433
Högbom J. A., 1974, A&AS, 15, 417
Hunter J. D., 2007, Comput. Sci. Eng., 9, 90
Hurley-Walker N. et al., 2017, MNRAS, 464, 1146
Hurley-Walker N. et al., 2022, PASA, 39, e035
Ignesti A. et al., 2022, ApJ, 924, 64
Jelić V. et al., 2008, MNRAS, 389, 1319
Joye W. A., Mandel E., 2003, in Payne H. E., Jedrzejewski R. I., Hook R. N., eds, ASP Conf. Ser. Vol.295, Astronomical Data Analysis Software and Systems XII. Astron. Soc. Pac., San Francisco, p. 489
Junklewitz H., Bell M. R., Enßlin T., 2015, A&A, 581, A59
Kazemi S., Yatawatta S., Zaroubi S., Lampropoulos P., de Bruyn A. G., Koopmans L. V. E., Noordam J., 2011, MNRAS, 414, 1656
Li F., Cornwell T. J., de Hoog F., 2011, A&A, 528, A31
Line J. L. B. et al., 2020, PASA, 37, e027
Liu A., Shaw J. R., 2020, PASP, 132, 062001
Lonsdale C. J., Morison I., 1980, Nature, 288, 66
Mazumder A., Datta A., Chakraborty A., Majumdar S., 2022, MNRAS, 515, 4020
McKean J. P. et al., 2016, MNRAS, 463, 3143
McKinney W., 2010, in van der Walt S., Millman J., eds, Proceedings of the 9th Python in Science Conference. SciPy 2010, Austin, p. 56

Mertens F. G., Ghosh A., Koopmans L. V. E., 2018, MNRAS, 478, 3640
Mitchell D. A., Greenhill L. J., Wayth R. B., Sault R. J., Lonsdale C. J., Cappallo R. J., Morales M. F., Ord S. M., 2008, IEEE J. Sel. Top. Signal Process., 2, 707
Mohan N., Rafferty D., 2015, PyBDSF: Python Blob Detection and Source Finder, Astrophysics Source Code Library, preprint(ascl:1502.007)
Morabito L. K. et al., 2022, A&A, 658, A1
Offringa A. R., Smirnov O. M., 2017, MNRAS, 471, 301
Offringa A. R. et al., 2014, MNRAS, 444, 606
Offringa A. R. et al., 2016, MNRAS, 458, 1057
Pandey V. N., Koopmans L. V. E., Tiesinga E., Albers W., Koers H. U. A., 2020, in Pizzo R., Deul E. R., Mol J. D., de Plaa J., Verkouter H., eds, ASP Conf. Ser. Vol.527, Astronomical Data Analysis Software and Systems XXIX. Astron. Soc. Pac., San Francisco, p. 473
Pâris I. et al., 2014, A&A, 563, A54
Parsons A. R., Backer D. C., 2009, AJ, 138, 219
Perley R. A., Chandler C. J., Butler B. J., Wrobel J. M., 2011, ApJ, 739, L1
Pilkington J. D. H., Scott J. F., 1965, MmRAS, 69, 183
Rau U., Cornwell T. J., 2011, A&A, 532, A71
Rau U., Bhatnagar S., Owen F. N., 2016, AJ, 152, 124
Refregier A., 2003, MNRAS, 338, 35
Rich J. W., de Blok W. J. G., Cornwell T. J., Brinks E., Walter F., Bagetakos I., Kennicutt R. C., J., 2008, AJ, 136, 2897
Riding J. L., Mitchell D. A., Webster R. L., 2017, in Lorente N. P. F., Shortridge K., Wayth R., eds, ASP Conf. Ser. Vol.512, Astronomical Data Analysis Software and Systems XXV, Astron. Soc. Pac., San Francisco, p. 257
Sault R. J., Wieringa M. H., 1994, A&AS, 108, 585
Schwab F. R., 1984, AJ, 89, 1076
Shaver P. A., Windhorst R. A., Madau P., de Bruyn A. G., 1999, A&A, 345, 380
Shimwell T. W. et al., 2017, A&A, 598, A104
Shimwell T. W. et al., 2019, A&A, 622, A1
Shimwell T. W. et al., 2022, A&A, 659, A1
Smirnov O. M., 2011, A&A, 527, A106
Spinrad H., Djorgovski S., Marr J., Aguilar L., 1985, PASP, 97, 932
Starck J.-L., Fadili J., Murtagh F., 2007, IEEE Trans. Image Process., 16, 297
Tasse C., van der Tol S., van Zwieten J., van Diepen G., Bhatnagar S., 2013, A&A, 553, A105
Terris M., Dabbech A., Tang C., Wiaux Y., 2022, MNRAS, 518, 604
Thompson A. R., Moran J. M., Swenson G. W. J., 2017, Interferometry and Synthesis in Radio Astronomy, 3rdedn. Springer, Berlin
Timmerman R. et al., 2022a, A&A, 658, A5
Timmerman R., van Weeren R. J., Botteon A., Röttgering H. J. A., McNamara B. R., Sweijen F., Bîrzan L., Morabito L. K., 2022b, A&A, 668, A65
Virtanen P. et al., 2020, Nat. Methods, 17, 261
Vollmer B., Davoust E., Dubois P., Genova F., Ochsenbein F., van Driel W., 2005, A&A, 431, 1177
Whyborn N. D., Browne I. W. A., Wilkinson P. N., Porcas R. W., Spinrad H., 1985, MNRAS, 214, 55
Wiaux Y., Jacques L., Puy G., Scaife A. M. M., Vandergheynst P., 2009, MNRAS, 395, 1733
Yatawatta S., 2011, in 2011 XXXth URSI General Assembly and Scientific Symposium. IEEE, Piscataway, p. 1–4
Yatawatta S., 2015, MNRAS, 449, 4506
Yatawatta S., 2016, in Proceedings of the 24th European Signal Processing Conference (EUSIPCO-2016) . IEEE, Piscataway, p. 265–269
Yatawatta S. et al., 2013, A&A, 550, A136
Yatawatta S., Diblen F., Spreeuw H., 2017, in 2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP). IEEE, Piscataway, p. 1–5
Ye H., Gull S. F., Tan S. M., Nikolic B., 2022, MNRAS, 510, 4110
de Gasperin F., Intema H. T., Frail D. A., 2018, MNRAS, 474, 5008
van Diepen G., Dijkema T. J., Offringa A., 2018, DPPP: Default Pre-Processing Pipeline, Astrophysics Source Code Library, preprint(ascl:1804.003)

Table A1. Four lines from the clean component list file after the first forced-spectrum fitting. From top to bottom: point component from the core region, and Gaussian components from lobe S, lobe N, and from the surrounding regions. The original file provided also the reference frequency at which the flux density *I* (i.e. *S* with the notation adopted in this paper) is estimated, which is 144.30 MHz for all the components, and the orientation of the Gaussians, which is always zero because only circular Gaussians are modelled by the MS deconvolution.

Name	Type	RA	Dec	I (Jy)	Spectral index	Major axis (arcsec)	Minor axis (arcsec)	Orientation (deg)
0c1069	POINT	$08^h34^m54\overset{s}{.}906$	$55°34'20\overset{''}{.}950$	0.293	1.738	–	–	–
s3c41	GAUSSIAN	$08^h34^m54\overset{s}{.}971$	$55°34'18\overset{''}{.}750$	0.010	-0.724	1.373	1.373	0
s4c30	GAUSSIAN	$08^h34^m54\overset{s}{.}567$	$55°34'24\overset{''}{.}950$	0.149	-0.897	2.746	2.746	0
s1c0	GAUSSIAN	$08^h34^m55\overset{s}{.}080$	$55°34'12\overset{''}{.}375$	0.002	-1.018	0.343	0.343	0

van Haarlem M. P. et al., 2013, A&A, 556, A2

APPENDIX A: WSCLEAN MODEL CATALOGUE

In Table A1, we report four example components from the clean component list file that WSCLEAN produces, one line for each of the regions into which we have divided 4C+55.16 (see Section 3.2). More details about this file format can be found at https://wsclean.readthedocs.io/en/latest/component_list.html.

APPENDIX B: SPECTRAL INDICES OBSERVED IN PIXEL-BASED MODELS

Here, we derive the spectral index that is observed in a model image when components overlap. The situation is described in Fig. 1. We consider a Gaussian component with the peak in position A and a point component in position B, whose spectra are described by a power law, with spectral indices α_A and α_B, respectively, constant over their full shape. The flux density of the Gaussian is

$$S_G(\nu) = S_G(\nu_0) \left(\frac{\nu}{\nu_0} \right)^{\alpha_A} , \qquad (B1)$$

where ν_0 is the reference frequency at which the normalization is evaluated, and similarly for the point source S_P, with spectral index α_B. Since $\alpha_A \neq \alpha_B$, there is a frequency ν_{int} where the two spectra intersect; without loss of generality, we can choose to normalize the flux density at $\nu_0 = \nu_{int}$, so that $S_G(\nu_0) = S_P(\nu_0)$ in B, where the components coincide.

When the spectral index of the resulting (gridded) model image is observed, the resulting flux density in B is the sum of $S_G(\nu)$ and $S_P(\nu)$, such that

$$S(\nu) = S_P(\nu_0) \left[\left(\frac{\nu}{\nu_0} \right)^{\alpha_A} + \left(\frac{\nu}{\nu_0} \right)^{\alpha_B} \right] , \qquad (B2)$$

which is not a power law, as shown in the bottom panel of Fig. 1. However, when we analyse the produced spectral indices (as in the first row of Fig. 4), we fit a power law pixel-by-pixel: in this case, the resulting spectral index α is a combination of α_A and α_B and changes with the frequency range over which the fit is performed.

To better understand what is the expected α in B, we move to a log–log space, so that the flux density is now defined as

$$\log[S(\nu)] = \log[S_P(\nu_0)] + \log\left[\left(\frac{\nu}{\nu_0} \right)^{\alpha_A} + \left(\frac{\nu}{\nu_0} \right)^{\alpha_B} \right] . \qquad (B3)$$

Then, the spectral index – or the slope – of $S(\nu)$ can be calculated by the derivative of $\log S$ with respect to $\log \nu$

$$\alpha(\nu) = \frac{d \log S}{d \log \nu} = \frac{\alpha_A \nu_0^{\alpha_B} \nu^{\alpha_A} + \alpha_B \nu_0^{\alpha_A} \nu^{\alpha_B}}{\nu_0^{\alpha_B} \nu^{\alpha_A} + \nu_0^{\alpha_A} \nu^{\alpha_B}} , \qquad (B4)$$

which is a function of ν, as expected.

With $\alpha_A > \alpha_B$, we can consider three special cases: (1) if $\nu \ll \nu_0$, we can rewrite equation (B4) as

$$\alpha(\nu) = \frac{\alpha_A (\nu/\nu_0)^{\alpha_A - \alpha_B} + \alpha_B}{(\nu/\nu_0)^{\alpha_A - \alpha_B} + 1} , \qquad (B5)$$

from which, we find $\alpha \approx \alpha_B$; (2) if $\nu \gg \nu_0$, we can rewrite equation (B4) as

$$\alpha(\nu) = \frac{\alpha_A + \alpha_B (\nu/\nu_0)^{\alpha_B - \alpha_A}}{(\nu/\nu_0)^{\alpha_B - \alpha_A} + 1} , \qquad (B6)$$

from which we find $\alpha \approx \alpha_A$; (3) if $\nu \approx \nu_0$, we obtain

$$\alpha \approx \frac{\alpha_A + \alpha_B}{2} . \qquad (B7)$$

This demonstrates that the slope of the sum of two power laws is $\alpha_B \lesssim \alpha(\nu) \lesssim \alpha_A$.

APPENDIX C: SELF-CALIBRATION OF 4C+55.16 DATA

Here, we provide a more detailed description of the self-calibration process that was performed to improve the quality of the 4C+55.16 data, in order to meet the requirements for testing the forced-spectrum method. The models used for this operation are obtained from imaging the data with WSCLEAN.

We start splitting the full bandwidth of 45 MHz into 12 output channels, each ∼4 MHz wide, which are jointly cleaned with no MF weighting. Using uniform weighting results in an integrated synthesized beam with an FWHM of 207×143 mas. The smallest synthesized beam is achieved for the image in the frequency range 161–165 MHz, with an FWHM of 128×184 mas. To produce shallow Stokes I images, we use a pixel scale of 25×25 mas and the W-gridding algorithm, which is more accurate and – in our case – faster than the standard W-stacking method of WSCLEAN (Arras et al. 2021; Ye et al. 2022). We clean down to an initial threshold of 50σ and a final threshold of 10σ, where $\sigma = 90\,\mu$Jy beam^{-1} for the frequency-integrated image. The automasking algorithm was used to generate scale-dependent masks that were used to constrain the cleaning. Using such high thresholds ensures that no artefacts are included in the models and that visibilities are calibrated by real components only. The obtained model is directly saved in the measurement set file of the data.

The calibration step is then performed with DP3. We use the output models within the `gaincal` algorithm to solve only the diagonal phases of the Jones matrix (i.e. a 2×2 complex matrix for every station) for each channel and 4 time intervals (i.e. one solution every 64 s). These solutions are then applied to the data and a new imaging step is started.

In the next self-calibration iterations, we gradually decrease the imaging thresholds, reaching 12σ and 3σ for the initial and the

final thresholds. This is possible because the image quality improves after each iteration, going from an initial image dynamic range of 23 000 to 30 000 after the last iteration, where we obtain the deepest integrated image, with a noise of $67\,\mu\mathrm{Jy\,beam^{-1}}$. Both the image noise and the dynamic range are always evaluated from imaging 12 output channels with an initial threshold of 8σ and a final threshold of 3σ through all the self-calibration step, to make the comparison as fair as possible. In addition to decreasing the cleaning thresholds, also the solution interval is reduced from 4 time-steps to 2 during the second self-calibration iteration, and then to 1 (i.e. one solution every 16 s) from the third iteration to the end. However, solving with a model every 4 MHz does not completely solve the phase problem that causes artefacts around the core, because a single solution is applied to all the channels within that frequency range. In fact, issues are still present when every frequency channel is imaged. For this reason, at a certain iteration of the self-calibration, we start imaging every

channel individually, obtaining 92 output images. Using 92 channels with no MF weighting degrades the synthesized beam: the integrated image now has a beam with an FWHM of 255×184 mas, while the best synthesized beam is achieved in the range $163 - 163.5\,\mathrm{MHz}$ with an FWHM of 166×225 mas.

After having solved the calibration issue in the core structure, we perform a couple more iterations solving diagonal gains (i.e. both amplitude and phase in the diagonal of the Jones matrix), to make sure that the spatially integrated spectrum of the source follows the expected slope. To do this, we rescale the output models to a flux density of 8.66 Jy at 142.67 MHz using a power law with $\alpha = -0.02$ (see Section 3.2). This model is then predicted into the measurement set file and used to evaluate the gains.

This paper has been typeset from a TeX/LaTeX file prepared by the author.