

# A method for determining radio continuum spectra and its application to large surveys

B. Vollmer<sup>1,2</sup>, E. Davoust<sup>3</sup>, P. Dubois<sup>1</sup>, F. Genova<sup>1</sup>, F. Ochsenbein<sup>1</sup>, and W. van Driel<sup>4</sup>

<sup>1</sup> CDS, Observatoire astronomique de Strasbourg, UMR 7550, 11 rue de l'Université, 67000 Strasbourg, France  
e-mail: bvollmer@astro.u-strasbg.fr

<sup>2</sup> Max-Planck-Institut für Radioastronomie, Auf dem Hügel 69, 53121 Bonn Germany

<sup>3</sup> UMR 5572, Observatoire Midi-Pyrénées, 14 avenue E. Belin, 31400 Toulouse, France

<sup>4</sup> Observatoire de Paris, Section de Meudon, GEPI, CNRS UMR 8111 and Université Paris 7, 5 place Jules Janssen, 92195 Meudon Cedex, France

Received 30 March 2004 / Accepted 19 October 2004

**Abstract.** A new tool to extract cross-identifications and radio continuum spectra from radio catalogues contained in the VIZIER database of the CDS is presented. The code can handle radio surveys at different frequencies with different resolutions. It has been applied to 22 survey catalogues at 11 different frequencies containing a total of 3.5 million sources, which resulted in over 700 000 independent radio cross-identifications and ~67 000 independent radio spectra with more than two frequency points. A validation of the code has been performed using independent radio cross-correlations from the literature. The mean error of the determined spectral index is  $\pm 0.3$ . The code produces an output of variable format that can easily be adapted to the purpose of the user.

**Key words.** astronomical data bases: miscellaneous – radio continuum: general

## 1. Introduction

The total number of records in radio-source catalogues has increased dramatically in the last two decades. Three major increases are due to R. Dixon's "Master Source List" in the seventies (for the first version see Dixon 1970; for an error list see Andernach 1989), to the 87GB (Gregory & Condon 1991), GB6 (Gregory et al. 1996) and PMN (Wright et al. 1994, 1996; Griffith et al. 1994, 1995) surveys in 1991 and to the release of NVSS (Condon et al. 1998), FIRST (White et al. 1998) and WENSS (Rengelink et al. 1997) in 1996/1997 (see also Andernach 1999). To study the nature of these celestial objects detected at radio frequencies and to take full advantage of the huge amount of information contained in the catalogues, one has to know their spectral energy distribution (SED) over as large a frequency range as available. Since each catalogue (except Dixon's master source list) contains information at a single frequency, cross-identification between different catalogues is essential for the study of these sources.

Radio source cross-identifications in the centimeter to meter wavelength domain are particularly difficult to obtain, because the underlying radio surveys can have huge differences in sensitivity and/or spatial resolution. Since the resolution depends on the observed frequency and the telescope diameter, low frequency surveys made with a small single-dish telescope can have resolutions of up to tens of arcmins, while high-frequency observations with a large single dish telescope or an

interferometer can have resolutions of a few arcsecs to tens of arcsecs.

On the other hand, the cross-identification of radio sources at different frequencies is made easier by using the fact that, in the vast majority of sources, the SED has a power-law distribution. The radiation mechanism is either synchrotron emission from relativistic electrons gyrating in a magnetic field, or emission of hot thermal electrons. Synchrotron emission produces a power law spectrum with a possible cut-off or reversal of the spectral index at low frequencies due to self-absorption or Comptonization. The spectrum of thermal electrons is flat, at least in the optically thin domain. Over the frequency range in which the majority of radio surveys were made, the spectra are thus well defined by a power law, i.e. as a straight line in the log (flux density)–log (frequency) plot commonly used in radio continuum astronomy.

The cross-identification procedure assigns radio sources at different frequencies to one physical object. In this way accurate radio positions can be determined for these objects using the high frequency observations, their radiation processes can be studied, and a search for specific objects (data mining) becomes possible.

The VIZIER database<sup>1</sup> (Ochsenbein et al. 2000) at the Centre de Données astronomique de Strasbourg (CDS) contains approximately 500 catalogues with radio data.

<sup>1</sup> <http://vizier.u-strasbg.fr/viz-bin/VizieR>

Of these 500, about 70 catalogues are from systematic surveys (for a list of the major surveys see Andernach 1999). Using VIZIER, only a cone search (where a central position and a radius is used) is possible on all, or on a subset of, these catalogues. This procedure gives a list of all radio sources within the search region without any cross-identification.

Cross-identifications of radio sources within SIMBAD<sup>2</sup> (Wenger et al. 2000) are essentially made on bibliographic grounds. Sources are only merged when a newly published radio catalogue gives alternate names, which were already known to SIMBAD. Thus there is a clear need to establish links between the radio catalogues (cross-identifications) and to include the sources in SIMBAD.

### 1.1. Radio surveys

There are two main kinds of radio surveys: (i) single dish; and (ii) interferometric. Since interferometers cannot detect extended structures larger than the size corresponding to the angular resolution of the shortest baselines (ten/tens of arcmin for compact configurations), only structures of this size, or smaller, can be detected in sources of large extent. This makes the identification and unique spectral index determination complex for very extended sources, if they were observed with both single-dish and interferometer telescopes.

In addition, the observations can be divided into two subclasses: (i) systematic surveys; and (ii) surveys of a given source sample (e.g. SN remnants, HII regions, etc.). While the first category is more suitable for integration into VIZIER, the second one is easier to feed into SIMBAD. Out of the 500 catalogues listed in VIZIER when choosing “radio”, about 220 contain independent observations. Of these, about 80 represent systematic surveys and about 140 surveys of source samples.

The characteristics of radio surveys that have to be taken into account when searching for cross-identifications are

- frequency;
- sky coverage;
- angular resolution (half power beam width, HPBW);
- sensitivity to point-like and extended sources.

From these surveys, source catalogues are established. The sources are extracted from 2D maps using 2D Gaussian fits, which have in principle 4 parameters: (i) the center position; (ii) the major axis; (iii) the minor axis; and (iv) the position angle. One distinguishes “map” parameters and “sky” parameters of sources. “Map” parameters are the extent along the major and minor axis from the Gaussian fit, whereas “sky” parameters are the deconvolved ones. If  $\Theta$  is the survey resolution (HPBW) and  $d$  the source extent on the map, then the true source extent is  $d_{\text{true}} = \sqrt{d^2 - \Theta^2}$ .

The catalogues can have different formats. For a given source one may find positions, position errors, a name, the peak and integrated flux densities and their errors, the major/minor axis, the position angle and various flags (confusion, extended source, warnings, etc.). The use of unified content

descriptors (UCDs<sup>3</sup>), which is the classification scheme in which all the astronomical parameters accessible in VIZIER are stored, will make the data access much easier in the future.

### 1.2. Cross-identification of radio sources

There are three categories of cross-identification between radio sources:

- cross-identification based on the proximity between two sources;
- cross-identification based on the radio spectrum;
- cross-identification based on other physical characteristics (e.g. galaxies, SN remnants, quasars, AGNs etc.).

The last two kinds of cross-identifications are called “value-added” cross-identification. In the present paper we present a tool for cross-identification called SPECIFIND which uses a proximity criterion and assumes that all sources have a power law radio spectrum.

This paper is structured in the following way: Sect. 2 describes the method used to make the catalogue tables uniform, which is required because the catalogue table entries often differ. The code algorithm is discussed in Sect. 3, followed by a discussion on the detailed code structure (Sect. 4). The code performance is presented in Sect. 5 and the results are shown in Sect. 6. The code is validated by the comparison of our spectral indices with independent estimates from the literature (Sect. 7). The summary and conclusions are given in Sect. 8.

## 2. Preparation and uniformisation of radio catalogues

The main difficulty in treating simultaneously different radio catalogue tables lies in the variety of their table structures. All tables have entries for the name, coordinates, and the flux density of the sources. The name and coordinates can be in different epochs (B1950 or J2000). Other possible entries are:

- position errors in right ascension and declination (including pointing errors and the uncertainty with which the center of the 2D Gaussian could be determined);
- flux density error;
- major/minor axis diameter of the source of the fitted ellipse (not deconvolved);
- position angle of the fitted ellipse;
- different sorts of flags (warning, border of the observed field, possible confusion, interferences, etc.).

This variety of entries makes a uniformisation of the radio catalogues unavoidable.

SPECIFIND uses its own standard for the catalogue entries. All coordinates are in J2000 and the source names are chosen to be in accordance with the SIMBAD nomenclature. Table 1 shows the list of SPECIFIND catalogue entries. The integrated flux densities  $S_{\text{int}}$  are taken directly from the radio catalogues, except for the GB6, 87GB and MIYUN catalogues, which give

<sup>2</sup> <http://simbad.u-strasbg.fr/sim-fid.pl>

<sup>3</sup> See <http://vizier.u-strasbg.fr/doc/UCD/inVizieR.htm>

**Table 1.** SPECIFIND catalogue standard.

RA(2000)	arcsec		REAL
RA_error	arcsec		REAL
Dec(2000)	arcsec		REAL
Dec_error	arcsec		REAL
Flux	mJy		REAL
flux_error	mJy		REAL
Wflag	Warning	0/1	INT
Cflag	Confused source	0/1	INT
Eflag	Extended source	0/1	INT
Majaxis	Major axis (units of beamsize)		REAL
Minaxis	Minor axis (units of beamsize)		REAL
PA	Position angle (degrees)		REAL
name2000	Source name J2000		CHAR
name1950	Source name B1950		CHAR

**Table 2.** Definitions of the flux density error when not taken directly from the catalogues.

JVAS	$0.1 \times \text{Flux}$
BWE	$\sqrt{36. + (0.01 \times \text{Flux}^2)}$
PKS	Flux/SN
PKS	$0.2 \times \text{Flux}$
F3R	error code
B2	$0.2 \times \text{Flux}$
B3	$0.2 \times \text{Flux}$
TXS	$0.2 \times \text{Flux}$
WISH	flux_error or $0.15 \times \text{Flux}$
WENSS	flux_error or $0.15 \times \text{Flux}$
MIYUN	peak flux or Flux/SN
4C	$0.3 \times \text{Flux}$
4C	$0.3 \times \text{Flux}$

only peak flux densities  $S_{\text{peak}}$ . In the GB6 and 87GB catalogues we take the peak flux density as the integrated flux density for sources smaller than 1.1 times the beamsize (3'5), and calculate the integrated flux density as

$$S_{\text{int}} = S_{\text{peak}} \left( \frac{d_{\text{maj}}}{\Theta} \right) \left( \frac{d_{\text{min}}}{\Theta} \right), \quad (1)$$

for larger sources, where  $d_{\text{maj}}$  and  $d_{\text{min}}$  are the extents along the major and minor axis on the map, respectively. For the MIYUN catalogue we take the peak flux density as integrated flux density when the latter is not given explicitly. The values of the flux errors that are not directly taken from the catalogues are listed in Table 2.

The flags are based on those taken from the catalogues. For the moment, however, SPECIFIND does not make use of the different flags.

### 3. SPECIFIND – the algorithm

SPECIFIND is a hierarchical code. It classifies a source  $j$  as parent, sibling or child with respect to a given source  $i$  at different stages where stage 2 and 3 are refinements of stage 1.

- Stage 1: depending on proximity criteria:
  - parent: source  $j$  has a larger extent or was observed with a lower angular resolution than source  $i$ ;
  - sibling: source  $j$  has a comparable extent or was observed with a comparable angular resolution (within 25%) to that of source  $i$ ;
  - child: source  $j$  has a smaller extent or was observed with a higher angular resolution than source  $i$ .
- Stage 2: depending on flux densities at the same frequency:
  - parent: source  $j$  has a larger extent or resolution and has a larger flux density than source  $i$ ;
  - sibling: source  $j$  has a comparable extent or resolution and has the same flux density within the errors as source  $i$ ;
  - child: source  $j$  has a smaller extent or resolution and a smaller flux density than source  $i$ .
- Stage 3: depending on flux densities at different frequencies, based on the expected radio spectral index:
  - parent: source  $j$  has a larger flux density than expected from the radio spectrum that includes source  $i$ ;
  - sibling: source  $j$  fits into the radio spectrum that includes source  $i$ ;
  - child: source  $j$  has a smaller flux density than expected from the radio spectrum that includes source  $i$ .

At the end of this procedure source  $i$  and its siblings are considered the same source. If there are parents, source  $i$  might be a resolved sub-source of source  $j$ . If there are children, source  $i$  might be extended without resolving the children, and the children represent the sub-sources of source  $i$ . SPECIFIND also adds the flux densities of the children of source  $i$  at the same frequency. If the sum equals the flux density of source  $i$  within the errors, then source  $i$  is considered as extended without resolving the children.

While the parents and children at the same frequency are not used for the moment, the parents and children at different frequencies are taken together with the siblings to determine the radio spectrum.

### 4. SPECIFIND – detailed code structure

#### 4.1. Proximity search

The proximity search is done using the treecode-method routines written by J. Barnes (see e.g. Barnes & Hut 1986<sup>4</sup>). Since this code is well documented, we will not describe the method nor the routines. We adapted these routines to spherical geometry. The angular distance between two sources, which are located at  $(\alpha_1, \delta_1)$  and  $(\alpha_2, \delta_2)$  is calculated in spherical geometry:

$$d = \arccos(\cos(\alpha_1 - \alpha_2) \cos(\delta_1) \cos(\delta_2) + \sin(\delta_1) \sin(\delta_2)). \quad (2)$$

<sup>4</sup> <http://www.ifa.hawaii.edu/~barnes/software.html>

We included the possibility to check if source  $j$  is located within the Gaussian ellipsoid characterising source  $i$ . If this is the case, the separation of the sources into parents/siblings/children is done in the following way: let  $d_i = \max(\text{resolution}(i), \text{Majaxis}(i))$  and  $d_j = \max(\text{resolution}(j), \text{Majaxis}(j))$ , i.e. the maximum between the resolution with which source  $i/j$  was observed and their major axes. If  $d_j > 1.25 d_i$ , then source  $j$  is considered as a parent. If  $0.75 d_i \leq d_j \leq 1.25 d_i$ , source  $j$  is considered as a sibling. If  $d_j < 0.75 d_i$ , source  $j$  is considered as a child. We do not include a check for the positional error ellipsoid, because when they are given in a catalogue these are only small fractions of the beamsize. On the other hand, since we take the FWHM of the Gaussian fit as the source extent, it is not necessary to take into account the error ellipsoid of the fit.

Since the treecode works in plane geometry, the polar caps ( $|\delta| > 70^\circ$ ) have to be treated separately. The rest of the sky is divided into equal RA slices, which ensures an approximately equal number of sources per slice. For the next neighbour search within each slice, sources within a somewhat larger field than the slice are used. The RA offset between this field and the slice is taken to be three times the largest source extent of all catalogues.

Since the largest extent of all sources is  $1^\circ$ , the maximum overlap is  $3^\circ$ . At a declination of  $70^\circ$ , a linear separation in RA of  $(\alpha_1 - \alpha_2) = 3^\circ$  corresponds to an angular separation of  $1^\circ$ . Proximity searches within both polar regions were also performed without using the treecode ( $N \times N$  calculations), which showed no change of the result with respect to the treecode.

To check if source  $j$  is located within the ellipsoidal source extent of source  $i$ , the ARC<sup>5</sup> projection is used for all catalogues. This projection, which preserves angular distances, is used for Schmidt plates (to first order) and in mapping with single-dish radio telescopes. Although the SIN<sup>4</sup> projection is applied for interferometric maps, we used the ARC projection for these as well, since (i) the majority of our catalogues are single-dish measurements; (ii) the largest source extents are found in single dish observations because of their larger beamsize; and (iii) the ARC and SIN projections are similar. An additional search is included for sources that are located around RA = 00<sup>h</sup>00<sup>m</sup>00<sup>s</sup>.

#### 4.2. Joining sources at the same frequency

This routine takes into account the flux densities of the sources at the same frequency. Since the sources are frequently point-like, i.e. their size fitted from the map equals the beamsize/resolution of the antenna used (unresolved sources), a source observed with a given beamsize is classified as a parent of another source observed with a smaller beamsize at the same frequency. Consequently, a source observed with a given beamsize is classified as a child of another source observed with a larger beamsize at the same frequency. A source  $i$  can have only one parent but several children from a different survey at the same frequency. Some or all children may be the same physical

object – this can be verified with the help of the flux densities measured at the same frequency. The routine makes the following classification: a parent has a larger flux density and a larger extent/resolution, a sibling has an equal flux density within the errors and an equal extent/resolution within 25%, a child has a smaller flux density and a smaller extent/resolution.

#### 4.3. Check for family dependences

Once the sources are joined in the way described in Sect. 4.2, the family dependences are verified, i.e. for a given source cross-checks are performed. These checks are performed for all source entries;

- If source  $j$  is a sibling of source  $i$ , then source  $i$  must also be a sibling of source  $j$ .
- If source  $j$  is a child of source  $i$ , source  $i$  must be a parent of source  $j$ .
- If source  $j$  is a parent of source  $i$ , source  $i$  must be a child of source  $j$ .

In addition, the flux densities of all children are added to investigate whether the source has multiple components at a given frequency. If the sum of the flux densities of all siblings equals that of the source within the errors, it is considered as resolved and an internal flag is set, which tells the user at which frequency the source is resolved. It happens frequently that a non-Gaussian emission distribution consists of multiple components within the area of one beamsize when observed with a smaller beam.

#### 4.4. Spectrum-finding algorithm

This is the most important routine and thus the heart of SPECIFIND. It uses the method of the least absolute deviation to make a linear fit in the  $\log \nu - \log S_\nu$  plane, where  $\nu$  is the frequency and  $S_\nu$  is the flux density at frequency  $\nu$ . This method is more robust against outlying points in a spectrum than a standard least-squares deviation ( $\chi^2$ ) fit (see Press et al. 2002).

For this algorithm, the best way to find a maximum number of spectra without a too high risk of misidentifications is to set the flux density errors of all sources that are smaller than 20% of their flux density to this 20% value and multiplying flux density errors by a factor 1.5. In this way all catalogues have approximately the same relative error. Moreover, these relatively large errors can compensate for a not too strong flattening of the spectral index at low frequencies due to the synchrotron turnover or at high frequencies due to an increasing fraction of thermal emission.

The structure of the spectrum-finding algorithm is the following: for a given set of sources for which all family relations were determined, their flux measurements at different frequencies are grouped together into an array and sorted by frequency. If the number of different frequencies is greater than two, the spectrum-finding algorithm passes through the following steps:

1. a least absolute deviation fit in the  $\log S_\nu - \log \nu$  plane is performed:

$$\log S_\nu = \alpha \log \nu + \gamma; \quad (3)$$

<sup>5</sup> See Greisen (1994)

2. if the spectrum is determined more than once:  
if the number of sources that fit into the spectrum decreases, the old fit parameters are used;
3. check for sources that fit into the spectrum; if all sources fit, go to step 6;
4. if there are two sources of the same frequency, the one with the largest deviation from the fit is flagged and removed;
5. if all sources have different frequencies the source with the largest deviation from the fit is flagged and removed, go to step 1;
6. if there are more than two independent points left and if the ratio between the largest and the smallest frequency interval is greater than 0.02, make a final fit;
7. go to step 1 and make a second run with  $\alpha = -0.9$  and  $\gamma = \log S_\nu - \alpha \log \nu$  during the first  $N - 4$  steps of the loop, where  $N$  is the initial number of points in the spectrum ( $-0.9$  is the mean spectral index of all radio sources);
8. if the number of fitted points with fixed  $\gamma$  and  $\alpha$  exceeds that of the initial fitting procedure, this spectrum is accepted; otherwise the spectrum of the first fitting procedure is accepted.

Step 4 excludes variable sources that are observed at different epochs. If the majority of the points has a high angular resolution, extended sources that are resolved into sub-sources might also be discarded. This algorithm turned out to be the most promising for finding a maximum number of spectra with only a small risk of misidentifications (see Sect. 6). By definition it can only detect parts of a spectrum that follow a power law (see also Sect. 4.6). This algorithm is similar to that used by Verkhodanov et al. (2000) for the identification of radio spectra of decameter-wavelength sources.

In order to investigate the efficiency of the spectrum-finding algorithm, we inspected by eye the data of the sources for which no spectrum could be found. We optimised the code to find a maximum number of radio spectra with a relatively small number of misidentifications (see also Sect. 7).

#### 4.5. Check for frequency intervals and ambiguous sources

In order to avoid using points that are too close to each other in frequency, and therefore not independent, the frequency intervals between the different points of the spectrum are checked. The routine calculates the frequency intervals and determines the ratio between the second largest and the largest frequency interval. If this ratio is smaller than 0.02, the spectrum is rejected.

Then, in order to avoid ambiguous radio sources of a given frequency, which are attributed to two distinct physical objects, the “center of mass” coordinates are calculated for both objects, where the inverse of the survey resolution is used for the “mass”. The ambiguous source is then attributed solely to the object whose “center of mass” position is nearest to the source position.

#### 4.6. Completeness and uniqueness check for spectra

This routine ensures that if a source  $j$  fits the spectrum determined for source  $i$  (where source  $i$  is included), then source  $i$  also appears in the spectrum of source  $j$ . In this way it is ensured that a radio source belongs to only one single physical object.

In practice the spectrum-finding algorithm is too efficient. Through chance alignments, sources that were observed with a large beamsize or which have a large extent are sometimes identified as physically belonging together (via the radio spectrum). Thus it happens that, while a source  $j$  fits the spectrum determined for source  $i$ , source  $i$  is not included in the spectrum of source  $j$ . There may be several reasons for this: (i) the spectrum of source  $i$  is erroneous, (ii) the spectrum of source  $j$  is erroneous, (iii) the spectral index varies with frequency, (iv) the real errors on the flux densities of one of the sources are larger than 30%. The task to make all spectra consistent is quite complicated, because all sources are interconnected via their siblings, which are interconnected via their own siblings, etc.

In order to decide which spectrum to take, in case of inconsistency between two spectra, the following scheme is applied: let  $\alpha_i$ ,  $\alpha_j$  be the spectral indices of source  $i$  and  $j$ .

- (i) If the difference between the spectral indices is smaller than 0.3 ( $|\alpha_i - \alpha_j| < 0.3$ ) (see Sect. 7) both spectra are real.
- (ii) If  $|\alpha_i - \alpha_j| \geq 0.3$  and the frequency intervals within which they are determined are only marginally overlapping (20%) both spectra are real. The spectrum is thus approximated by two different slopes within two different frequency intervals.
- (iii) If (i) and (ii) are not the case, the spectrum with the larger number of independent frequency points is real. The other spectrum is discarded.
- (iv) If (i) and (ii) are not the case and numbers of independent frequency points are the same for both spectra:
  - (iva) if  $\alpha_i < 0$  and  $\alpha_j < 0$  and  $\alpha_i > \alpha_j$ , spectrum  $i$  is real and spectrum  $j$  is discarded;
  - (ivb) if  $\alpha_i < 0$  and  $\alpha_j < 0$  and  $\alpha_i < \alpha_j$ , spectrum  $j$  is real and spectrum  $i$  is discarded;
  - (ivc) if  $\alpha_i \times \alpha_j < 0$  and  $\alpha_i < 0$ , spectrum  $i$  is real and spectrum  $j$  is discarded;
  - (ivd) if  $\alpha_i \times \alpha_j < 0$  and  $\alpha_j < 0$ , spectrum  $j$  is real and spectrum  $i$  is discarded;
  - (ive) if  $\alpha_i > 0$  and  $\alpha_j > 0$  and  $\alpha_i < \alpha_j$ , spectrum  $i$  is real and spectrum  $j$  is discarded;
  - (ivf) if  $\alpha_i > 0$  and  $\alpha_j > 0$  and  $\alpha_i > \alpha_j$ , spectrum  $j$  is real and spectrum  $i$  is discarded. The spectra are only completed if the proximity criterion is satisfied, i.e. if the extents/beamsizes of the sources intersect.

The above-described procedure checks only for siblings of siblings. This does not take into account the case where the same source is attached as a sibling to two different sources without being a sibling of a sibling of any of these two sources. The only way to identify these cases is again to make a full proximity search using the treecode (Sect. 4.1). All nearby sources are then checked for common siblings. If the modulus of the difference between the spectral indices of the two sources is smaller than 0.3, the missing siblings are added to both sources if the proximity criterion is met (i.e., if the distance is smaller than half the sum of the source extents/resolutions). If the modulus of the difference between the spectral indices of the two sources is  $\geq 0.3$ , the siblings of sources  $i$  and  $j$  are modified in the following way: let the number of siblings of source  $i$  be greater than that of source  $j$ . If the numbers of siblings are the same, let the source with the steeper spectrum be source  $i$ . If source  $j$  is a sibling of source  $i$ , this sibling is removed. In addition, all siblings are removed from source  $j$  that have a different frequency than source  $j$ . If source  $j$  is not a sibling of source  $i$ , the common sibling is removed from source  $j$ .

In a final step the siblings of a given source are compared to the siblings of all its siblings. If there are less than three common siblings and both spectral indices are non zero and the numbers of siblings are different, the spectrum of the source with the smallest number of siblings is removed (i.e. the siblings at a different frequency than the source are removed). If this is not the case, both sources are complemented with the missing siblings.

#### 4.7. Output

At the end of the data processing for one subfield (RA slices and polar caps), the results are written in an ASCII file. Since all necessary information for all sources is stored in the code, the output format can be chosen freely and adapted to the user's purpose. For the moment we create two principal outputs: (i) a file with the information necessary to plot spectra; and (ii) a file that can be used as input for SIMBAD.

### 5. Code performance

On a PC with 512 MB RAM and a frequency of 1.4 GHz, the whole data processing of all 3.5 million sources can be performed in less than 3 h, less than one quarter of which is needed to read the input files and to write the final ASCII files. The proximity search in one single slice is performed in 5–10 min. This relatively long time is due to the spherical geometry.

### 6. Results

We have based our selection of radio source catalogues on a list of major surveys of discrete radio sources (see Table 1 of Andernach 1999)<sup>6</sup>. For the moment we have included the 22 largest of the 66 cited radio catalogues (Table 3). These catalogues were downloaded from the VIZIER database.

The 3C and 3CR catalogues were added for historical reasons. The GB6 and 87GB catalogues are not independent. In fact 87GB is based on a subset of the data used for GB6, which is more sensitive. The total number of sources is 3 488 352. The NVSS catalogue provides  $\sim 50\%$  of these entries. For this number of sources the sky is divided into 12 RA slices and the 2 polar caps ( $|\delta| > 70^\circ$ ).

After completion of the data processing, SPECIFIND found 757 894 independent associations, i.e. sources with at least one parent, sibling, or child. The number of independent spectra is 66 866, of which more than 90% include an NVSS point. SPECIFIND identified 5 objects with a spectral index  $\alpha < -2$ . Only pulsars and relic sources in galaxy clusters show such steep spectra. The inspection of these objects by eye showed that the objects, including a FIRST source, are extended (jet/lobe structure) and/or confused. The ensemble of 66 866 independent spectra forms the basis for our further analysis. The percentage of sources, for which spectra could be identified by SPECIFIND is listed in Col. 7 of Table 3.

The highest percentage of sources with detected spectra is 72% of all sources for the JVAS and B2 surveys. This is expected for the JVAS, because it is a survey of sources selected on the basis of their flat spectral index. The large surveys produce fewer spectra, because the other surveys do not cover the same region of the sky as deeply as the former. In the SUMSS survey only 1.8% of the source have spectra, because there is a lack of deep enough radio surveys in the southern hemisphere.

The distribution of the 66 866 spectral indices is shown in Fig. 1. The distribution peaks at  $\alpha \sim -0.9$ , which is consistent with previous works (Vigotti et al. 1989; Kulkarni et al. 1990; Zhang et al. 2003). There is a wing towards positive spectral indices, which is most probably caused by sources with a flat spectrum due to thermal electrons.

The number of sources as a function of the number of independent frequency points in the radio spectra for independent sources is shown in Fig. 2. About  $10^4$  sources have 5 points and about 50 sources have 8 points. The number of sources decreases approximately exponentially between 3 and 6 points and falls off more rapidly for an even larger number of points.

The spectral index as a function of the flux density at 325 MHz (49 cm; WENSS) is shown in Fig. 3. As already seen in Fig. 1, most of the spectral indices have values around  $-0.9$ . Since the NVSS and WENSS surveys are by far the deepest surveys that cover the largest area on the sky (the whole northern hemisphere), the largest number of associations is found for sources in these two catalogues. Zhang et al. (2003) have correlated them and found 185 800 corresponding sources, which represents the maximum number of sources for which a spectrum can be found. We found only half of them, because we require at least 3 points for the spectrum.

The straight, almost horizontal edge of the distribution in the left part of the plot (marked as (a) in Fig. 3) is due to a selection effect. For these low flux density sources with a steep spectrum, SPECIFIND found a source at 20 cm (NVSS) and 50 cm (WENSS), but none at 6 cm (GB6, BWE, 87GB, MITG, PMN), where the sensitivity of the surveys ( $\sim 20$  mJy) is insufficient. At higher flux densities ( $S_{325} > 300$  mJy) sources from other catalogues with lower sensitivities are found. Most

<sup>6</sup> Updated at [http://cats.sao.ru/doc/MAJOR\\_CATS.html](http://cats.sao.ru/doc/MAJOR_CATS.html)

**Table 3.** SPECIFIND catalogue entries.

Name	Interferometer (I) Single dish (S)	Frequency (MHz)	Resolution (arcmin)	$S_{\min}$ (mJy)	Number of sources	percentage of sources with identified spectrum	Reference
JVAS	I	8400	$5.5 \times 10^{-3}$	30	2246	72	(1)
GB6	S	4850	3.5	18	75 162	59	(2)
87GB	S	4850	3.5	25	54 579	64	(3)
BWE	S	4850	3.5	25	53 522	61	(4)
PMN	S	4850	3.5	20	50 814	30	(5)
MITG	S	4850	2.8	40	24 180	52	(6)
PKS	S	2700	8.0	50	8264	65	(7)
F3R	S	2700	4.3	40	6495	61	(8)
FIRST	I	1400	0.08333	1	811 117	1.5	(9)
NVSS	I	1400	0.75	2	1 773 484	3.6	(10)
WB	S	1400	10.	100	31 524	60	(11)
SUMSS	I	843	0.75	8	134 870	1.8	(12)
B2	I	408	8.0	250	9929	72	(13)
B3	I	408	5.0	100	13 340	66	(14)
MRC	I	408	3.0	700	12 141	73	(15)
TXS	I	365	0.1	250	66 841	57	(16)
WISH	I	325	0.9	10	90 357	8.4	(17)
WENSS	I	325	0.9	18	229 420	17	(18)
MIYUN	I	232	3.8	100	34 426	40	(19)
4C	I	178	11.5	2000	4844	53	(20)
3CR	I	178	6.0	5000	327	31	(21)
3C	I	159	10.0	7000	470	3.4	(22)

(1) Patnaik et al. (1992), Browne et al. (1998), Wilkinson et al. (1998).

(2) Gregory et al. (1996).

(3) Gregory & Condon (1991).

(4) Becker et al. (1991).

(5) Wright et al. (1994, 1996), Griffith et al. (1994, 1995).

(6) Bennett et al. (1986), Langston et al. (1990), Griffith et al. (1990, 1991).

(7) Otrupcek & Wright (1991).

(8) Fürst et al. (1990).

(9) White et al. (1998).

(10) Condon et al. (1998).

(11) White & Becker (1992).

(12) Mauch et al. (2003).

(13) Colla et al. (1970, 1972, 1973), Fanti et al. (1974).

(14) Ficarra et al. (1985).

(15) Douglas et al. (1996).

(16) Large et al. (1991).

(17) de Breuck et al. (2002).

(18) Rengelink et al. (1997).

(19) Zhang et al. (1997).

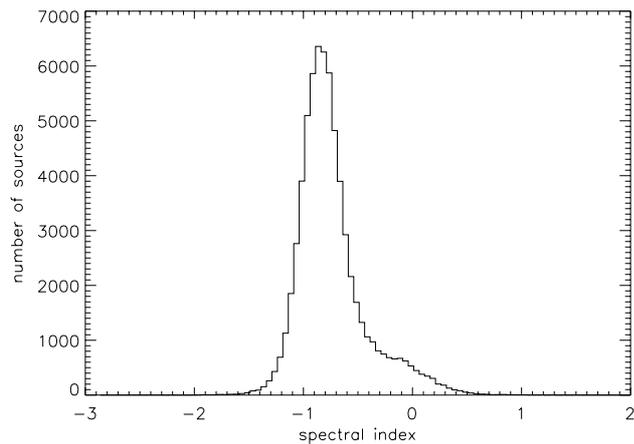
(20) Pilkington & Scott (1965), Gower et al. (1967).

(21) Bennett (1962).

(22) Edge et al. (1959).

of the sources have spectral indices  $< -0.6$  (Fig. 1). This translates into a limiting sensitivity of 100 mJy at 325 MHz. The number of WENSS sources with flux densities that exceed this value is 68 732, close to the number of identified radio

spectra. The vertical edge in the lower left part of the plot (marked as (b) in Fig. 3) is mainly due to the limiting flux density of the B3 survey.



**Fig. 1.** Number distribution of spectral indices.

If one wants to increase significantly the number of independent spectra with respect to the existing catalogues, a deep and extended survey at an independent wavelength, preferentially smaller than 10 cm, is needed.

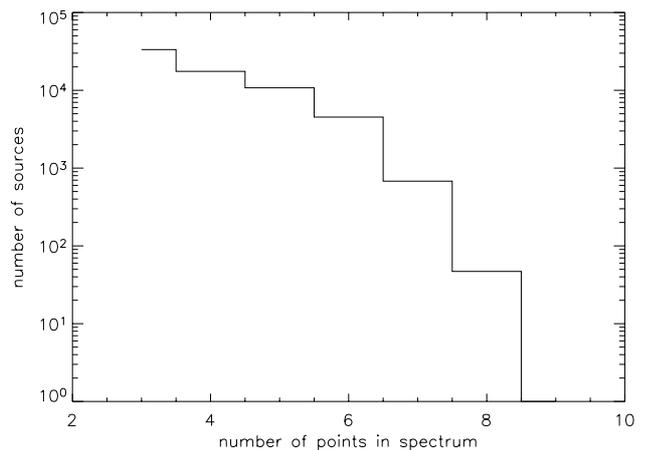
## 7. Code validation

In order to validate our algorithm, we compared the spectral indices determined by SPECIFIND with those given in various catalogues. The radio catalogues in the VIZIER database that include spectral indices are listed in Table 5; for the PMNS catalogue we calculated the spectral index from the data given in that catalogue. The columns are: (1) catalogue name; (2) the frequency used for the cross-identification in SPECIFIND; (3) other frequencies for the determination of the spectral index; (4) number of sources with spectral index. The catalogues B3, MITG, PKS, PMNS-S, and FA87 use only two frequencies for the determination of the spectral index, whereas the other catalogues use more than three frequencies. The percentage of sources for which a spectrum could be identified is listed in Col. 2 of Table 6. The high percentage of radio spectra obtained validates the SPECIFIND spectrum identification algorithm. For the direct comparison of spectral indices we only chose indices for which both methods used approximately the same frequency interval.

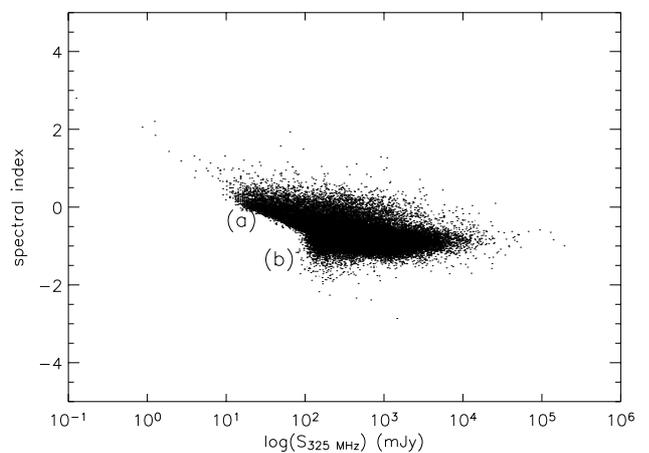
Only one pulsar spectrum could be identified by SPECIFIND (Table 4). This is due to the small flux densities and steep spectral indices of the pulsar population listed in PULSARS.

A comparison between the spectral indices given in these catalogues and those found by SPECIFIND are shown in Figs. 4–11.

In general, the scatter is smaller for smaller spectral indices, because there is no change in the spectral index in the frequency interval of interest. The increase of the scatter for large spectral indices (flatter spectra) in the AGN-QSO, FA87 and 1Jy catalogues (Figs. 6, 8 and 9) is due to the fact that these are radio sources whose spectra peak around 5 GHz. The determination of their spectral index therefore depends critically on the frequency interval used. These intervals are more precisely



**Fig. 2.** The number of sources as a function of the number of independent points in the radio spectra.



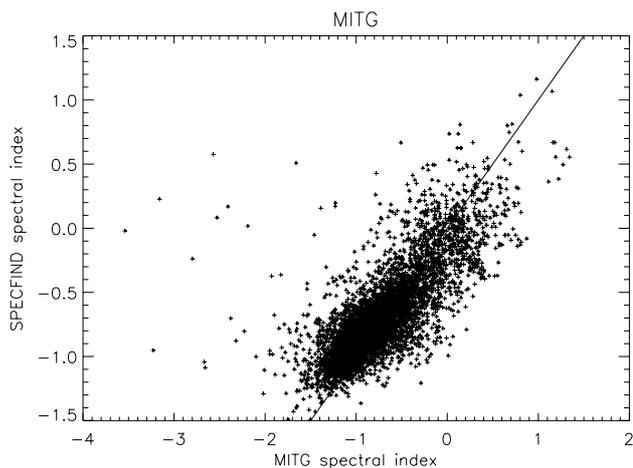
**Fig. 3.** The spectral index as a function of the flux density at 325 MHz.

**Table 4.** Pulsar spectrum identified by SPECIFIND.

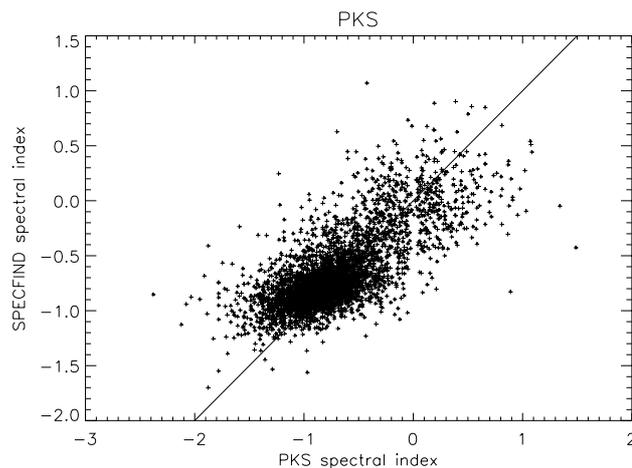
Source name	SI (SPECIFIND)	SI (PULSARS)
B0329 + 54	-1.7	-1.6

known for the other catalogues, which consequently show a smaller scatter.

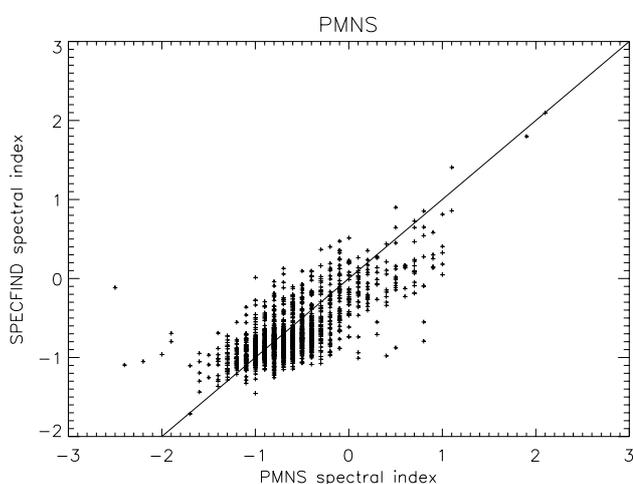
The standard deviations  $\sigma_{SI}$  of the difference between the spectral indices determined by SPECIFIND and those determined in the catalogues (Table 5) are listed in Table 6. The best correlation (0.10) between the two is found for the B3 catalogue, and the worst (0.36) for the AGN-QSO catalogue. The low consistency of spectral indices in the AGN-QSO catalogue may be either due to the small frequency range used (4850/2700 MHz), to variability, or both. In general, the standard deviation is  $\sim 0.3$ . Thus we conclude that the spectral indices determined by SPECIFIND have an error of about  $\pm 0.3$ .



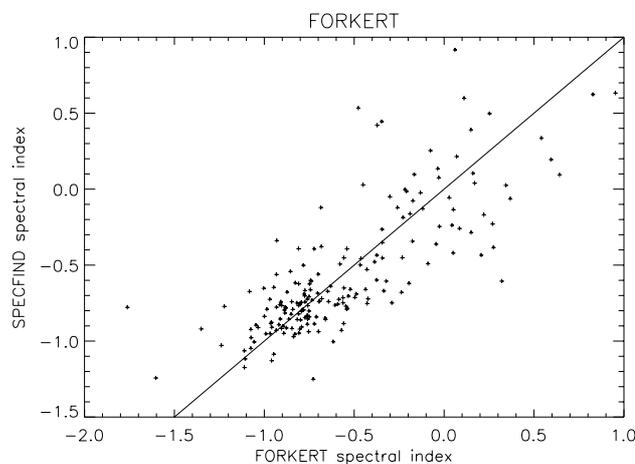
**Fig. 4.** MITG spectral index versus SPECFIND spectral index.



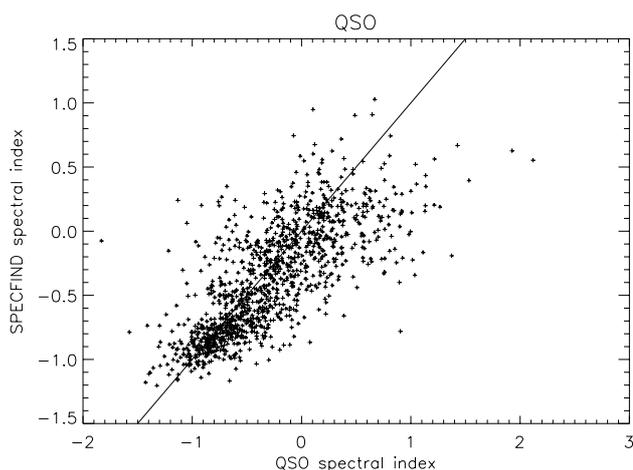
**Fig. 7.** PKS spectral index versus SPECFIND spectral index.



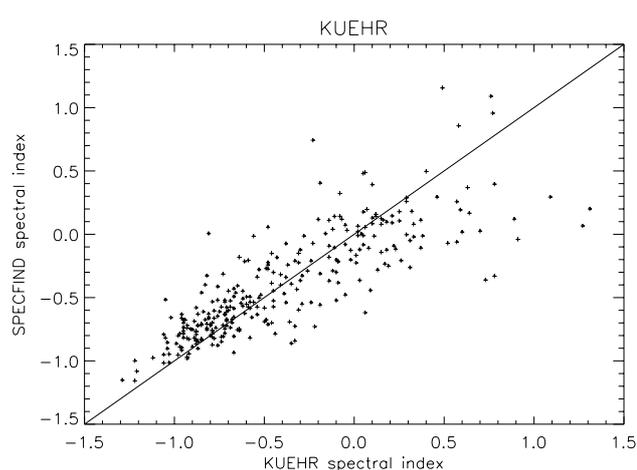
**Fig. 5.** PMN-S spectral index versus SPECFIND spectral index.



**Fig. 8.** FA87 spectral index versus SPECFIND spectral index.



**Fig. 6.** AGN-QSO spectral index versus SPECFIND spectral index.

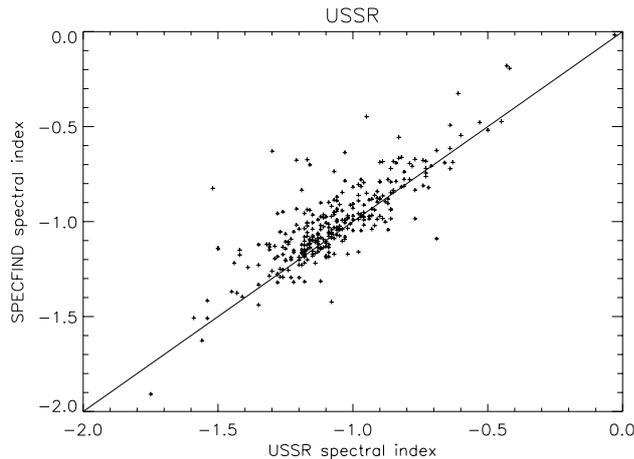


**Fig. 9.** 1Jy spectral index versus SPECFIND spectral index.

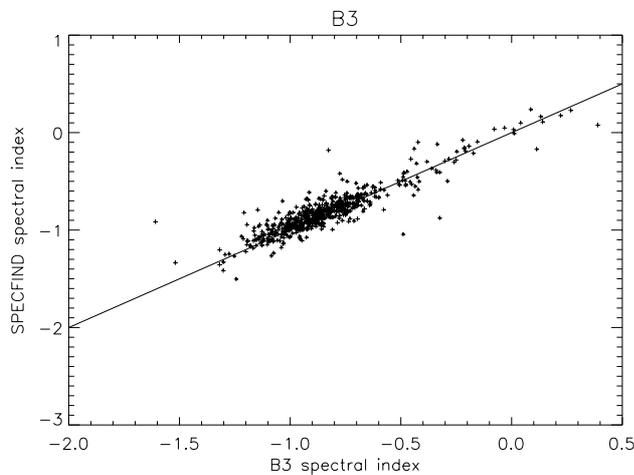
## 8. Conclusions

SPECFIND is a very efficient tool to identify radio spectra using radio catalogues of different formats. SPECFIND can handle radio surveys of very different resolutions and sensitivities. It has been applied to 22 survey catalogues at 11 different

frequencies containing a total of 3.5 million sources, leading to more than 700 000 independent radio cross-identifications and  $\sim 67\,000$  independent radio spectra with more than two independent frequencies. The code was tested and its results validated by a comparison between the spectral indices found by SPECFIND and those determined by other authors.



**Fig. 10.** USSR spectral index versus SPECIFIND spectral index.



**Fig. 11.** B3 spectral index versus SPECIFIND spectral index.

The determined spectral indices have an error of about  $\pm 0.3$ . Negative spectral indices have smaller errors, while the error of positive spectral indices can be larger, mainly because of the occurrence of a peak in the spectrum. The code is quite rapid (less than 3 h running time on a standard PC for 3.5 million sources) and since it is written in C, it can be run on virtually all PCs with at least 512 MB RAM. It produces an output of variable format that can be adapted easily to the purpose of the user. A special output to enter the cross-identifications into SIMBAD has been developed. The code has been optimised to find a maximum number of spectra with a relatively small number of misidentifications. It represents thus a promising tool to extract radio spectra from a large sample of radio catalogues, like those stored in VIZIER. Although at present only 22 catalogue entries can be accessed by SPECIFIND, this number will be increased in the future. The advantage of this procedure is that radio spectra and cross-identifications are established in advance, and can be stored in a “master list” within VIZIER, allowing a fast and efficient access. We expect to make the cross-identifications available soon via the VIZIER database.

*Acknowledgements.* We would like to thank H. Andernach for very helpful discussions and O. V. Verkhodanov for his comment on ambiguous sources and “centers of mass”.

**Table 5.** Catalogues with spectral indices.

Name	Frequency (MHz)	Frequency range (MHz)	Number of sources	Ref.
MITG	4850	365	7559	(1)
PMN-S	4850	2700	1847	(2)
AGN-QSO	4850	2700	1155	(3)
PKS	2700	5000	4150	(4)
FA87	2700	4750	211	(5)
1Jy	1400	2700–5000	299	(6)
USSR	1400	150–408	313	(7)
PULSARS	1400	300–1400	227	(8)
B3	408	1400	613	(9)

- (1) Drinkwater et al. (1997).
- (2) Wright et al. (1994).
- (3) Véron-Cetty & Véron (2003).
- (4) Bennett et al. (1986), Langston et al. (1990), Griffith et al. (1990, 1991).
- (5) Forkert & Altschuler (1987).
- (6) Kühn et al. (1981).
- (7) Roettgering et al. (1994).
- (8) Lorimer et al. (1995), Maron et al. (2000).
- (9) Kulkarni et al. (1990).

**Table 6.** Standard deviation of the spectral index difference.

Catalogue name	Percentage of sources with identified spectrum	$\sigma_{SI}$
MITG	100%	0.26
PMNS – S	100%	0.29
AGN – QSO	87%	0.36
PKS	100%	0.30
FA87	91%	0.28
1Jy	100%	0.29
USSR	99%	0.13
PULSARS	–	–
B3	100%	0.10

## References

- Andernach, H. 1989, *Bull. Inf. CDS*, 37, 139
- Andernach H. 1999, in *Internet Resources for Professional Astronomy, IXth Canary Islands Winter School of Astrophysics, Astrophysics with Large Databases in the Internet Age*, ed. M. Kidger, I. Perez-Fournon, & F. Sanchez (Cambridge: Cambridge Univ. Press), 67 [arXiv:astro-ph/9807346]
- Barnes, J., & Hut, P. 1986, *ApJ*, *Nature*, 324, 446
- Becker, R. H., White, R. L., & Edwards, A. L. 1991, *ApJS*, 75, 1 (BWE)
- Bennett, A. S. 1962, *MNRAS*, 68, 163 (3CR)
- Bennett, C. L., Lawrence, C. R., Burke, B. F., Hewitt, J. N., & Mahoney, J. 1986, *ApJS*, 61, 1 (MITG)
- Browne, I. W. A., Patnaik, A. R., Wilkinson, P. N., & Wrobel, J. M. 1998, *MNRAS*, 293, 257 (JVAS)
- Colla, G., Fanti, C., Fanti, R., et al. 1970, *A&AS*, 1, 281 (B2)
- Colla, G., Fanti, C., Fanti, R., et al. 1972, *A&AS*, 7, 1 (B2)
- Colla, G., Fanti, C., Fanti, R., et al. 1973, *A&AS*, 11, 291 (B2)

- Condon, J. J., Cotton, W. D., Greisen, E. W., et al. 1998, *AJ*, 115, 1693 (NVSS)
- Dixon, R. S. 1970, *ApJS*, 20, 1
- Drinkwater, M. J., Webster, R. L., Francis, P. J., et al. 1997, *MNRAS*, 284, 85
- de Breuck, C., Tang, Y., de Bruyn, A. G., Rottgering, H., & van Breugel, W. 2002, *A&A*, 394, 59 (WISH)
- Douglas, J. N., Bash, F. N., Bozayan, F. A., Torrence, G. W., & Wolfe, C. 1996, *AJ*, 111, 1945 (TXS)
- Edge, D. O., Shakeshaft, J. R., McAdam, W. B., Baldwin, J. E., & Archer, S. 1959, *MNRAS*, 68, 37 (3C)
- Fanti, C., Fanti, R., Ficarra, A., & Padrielli, L. 1974, *A&AS*, 18, 147 (B2)
- Ficarra, A., Grueff, G., & Tomassetti, G. 1985, *A&AS*, 59, 255 (B3)
- Forkert, T., & Altschuler, D. R. 1987, *A&AS*, 70, 77
- Fürst, E., Reich, W., Reich, P., & Reif, K. 1990, *A&AS*, 85, 805 (F3R)
- Gower, J. F. R., Scott, P. F., & Wills, D. 1967, *MNRAS*, 71, 49 (4C)
- Gregory, P. C., & Condon, J. J. 1991, *ApJS*, 75, 1011 (87GB)
- Gregory, P. C., Scott, W. K., Douglas, K., & Condon, J. J. 1996, *ApJS*, 103, 427 (GB6)
- Griffith, M., Langston, G., Heflin, M., et al. 1990, *ApJS*, 74, 129 (MITG)
- Griffith, M., Langston, G., Heflin, M., Conner, S., & Burke, B. 1991, *ApJS*, 75, 801 (MITG)
- Griffith, M. R., Wright, A. E., Burke, B. F., & Ekers, R. D. 1994, *ApJS*, 90, 179 (PMNT)
- Griffith, M. R., Wright, A. E., Burke, B. F., & Ekers, R. D. 1995, *ApJS*, 97, 347 (PMNE)
- Kulkarni, V. K., Mantovani, F., & Pauliny-Toth, I. I. 1990, *A&AS*, 82, 41
- Kühr, H., Witzel, A., Pauliny-Toth, I. I. K., & Nauber, U. 1981, *A&AS*, 45, 367
- Langston, G. I., Heflin, M. B., Conner, S. R., et al. 1990, *ApJS*, 72, 621 (MITG)
- Large, M. I., Cram, L. E., & Bruggess, A. M. 1991, *The Observatory*, 111, 72
- Lorimer, D. R., Yates, J. A., Lyne, A. G., & Gould, D. M. 1995, *MNRAS*, 273, 411
- Maron, O., Kijak, J., Kramer, M., & Wielebinski, R. 2000, *A&AS*, 147, 195
- Mauch, T., Murphy, T., Buttery, H. J., et al. 2003, *MNRAS*, 342, 1117
- Ochsenbein, F., Bauer, P., & Marcout, J. 2000, *A&AS*, 143, 23
- Otrupcek, R., & Wright, A. E. 1991, *PASAu*, 9, 1700 (PKS)
- Patnaik, A. R., Browne, I. W. A., Wilkinson, P. N., & Wrobel, J. M. 1992, *MNRAS*, 254, 655 (JVAS)
- Pilkington, J. D. H., & Scott, P. F. 1965, *MNRAS*, 69, 183 (4C)
- Press, W. H., Teukolsky, S. A., & Vetterling, W. T. 2002, *Numerical Recipes in C* (Cambridge: Cambridge Univ. Press)
- Rengelink, R. B., Tang, Y., de Bruyn, A. G., et al. 1997, *A&AS*, 124, 259
- Roettgering, H. J. A., Lacy, M., Miley, G. K., Chambers, K. C., & Saunders, R. 1994, *A&AS*, 108, 79
- Verkhodanov, O. V., Andernach, H., & Verkhodanova, N. V. 2000, *Bull. Spec. Astrophys. Obs.*, 49, 53 [arXiv:astro-ph/0008431]
- Véron-Cetty, M.-P., & Véron, P. 2003, *A&A*, 412, 399
- Vigotti, M., Grueff, G., Perley, R., Clark, B. G., & Bridle, A. H. 1989, *AJ*, 98, 419
- Wenger, M., Genova, F., Bonnarel, F., et al. 2000, *A&AS*, 143, 9
- White, R. L., & Becker, R. H. 1992, *ApJS*, 79, 331
- White, R. L., Becker, R. H., Helfand, D. J., & Gregg, M. D. 1998, *ApJ*, 475, 479 (FIRST)
- Wilkinson, P. N., Browne, I. W. A., Patnaik, A. R., Wrobel, J. M., & Sorathia, B. 1998, *MNRAS*, 300, 790 (JVAS)
- Wright, A. E., Griffith, M. R., Burke, B. F., & Ekers, R. D. 1994, *ApJS*, 91, 111 (PMNS)
- Wright, A. E., Griffith, M. R., Hunt, A. J., et al. 1996, *ApJS*, 103, 145 (PMNZ)
- Zhang, X., Zheng, Y., Chen, H., et al. 1997, *A&AS*, 121, 59 (MIYUN)
- Zhang, X., Reich, W., Reich, P., & Wielebinski, R. 2003, *A&A*, 404, 57