



**HAL**  
open science

## Toward Cosmicflows-4: The HI data catalog

A. Dupuy, H. M. Courtois, D. Guinet, R. B. Tully, E. Kourkchi

► **To cite this version:**

A. Dupuy, H. M. Courtois, D. Guinet, R. B. Tully, E. Kourkchi. Toward Cosmicflows-4: The HI data catalog. *Astronomy & Astrophysics - A&A*, 2021, 646, pp.A113. <10.1051/0004-6361/202039025>. <hal-03143631>

**HAL Id: hal-03143631**

**<https://hal.science/hal-03143631v1>**

Submitted on 16 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Toward Cosmicflows-4: The H<sub>I</sub> data catalog<sup>★</sup>

A. Dupuy<sup>1</sup>, H. M. Courtois<sup>1</sup>, D. Guinet<sup>1</sup>, R. B. Tully<sup>2</sup>, and E. Kourkchi<sup>2</sup>

<sup>1</sup> University of Lyon, UCB Lyon 1, CNRS/IN2P3, IUF, IP2I Lyon, UMR5822, 69622 Villeurbanne Cedex, France  
e-mail: dupuy@ipnl.in2p3.fr

<sup>2</sup> Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, HI 96822, USA

Received 24 July 2020 / Accepted 17 December 2020

## ABSTRACT

**Aims.** In this study, we present an update of a compilation of line width measurements of neutral atomic hydrogen (H<sub>I</sub>) galaxy spectra at 21 cm wavelength. Our All Digital HI (ADHI) catalog consists of the previous release augmented with our new HI observations and an analysis of archival data. This study provides the required HI information to measure the distances of spiral galaxies through the application of the Tully-Fisher (TF) relation.

**Methods.** We conducted observations at the Green Bank telescope (GBT) and reprocessed spectra obtained at the Nançay radiotelescope by the Nançay Interstellar Baryons Legacy Extragalactic Survey and Kinematics of the Local Universe collaborations and we analyzed the recently published full completion Arecibo Legacy Fast ALFA (ALFALFA) 100% survey in order to identify galaxies with good quality HI line width measurements.

**Results.** This paper adds new HI data adequate for TF use for 385 galaxies observed at GBT, 889 galaxies from archival Nançay spectra, and 1515 rescaled Arecibo ALFALFA spectra. In total, this release adds 1274 new good quality measurements to the ADHI catalog. Today, the ADHI database contains 18 874 galaxies, for which 15 433 have good quality data for TF use. The final goal is to compute accurate distances to spiral galaxies, which will be included in the next generation of peculiar velocities catalog: Cosmicflows-4.

**Key words.** methods: observational – line: profiles – catalogs – galaxies: distances and redshifts

## 1. Introduction

In 2006, we launched a large observational and theoretical program to study the large-scale structures and dynamics of the local Universe up to  $z \sim 0.1$ . Some remarkable outcomes of this project include the discovery of our home supercluster Laniakea (Tully et al. 2014), the Dipole Repeller (Hoffman et al. 2017, DR), the Cold Spot Repeller (Courtois et al. 2017), and the cartography of the Local Void (Tully et al. 2019).

Recently, we set up a distance-velocity calculator that is publicly accessible through the Extragalactic Distance Database<sup>1</sup> (EDD). It is for any astronomer needing information on the local gravitational velocity field at a specified location in the local Universe (Kourkchi et al. 2020a).

The HI line widths data in our Cosmicflows catalogs come from our observations at Green Bank telescope (GBT) and our reprocessing of spectra obtained at other giant radio telescope archives such as Parkes, Nançay, Effelsberg, and Arecibo, which were all gathered in the All Digital HI catalog (ADHI, Courtois et al. 2009, 2011; Courtois & Tully 2015). Table 1 shows a list of the various data and their respective publications in Col. 2, which was used to build the ADHI catalog. Column 1 corresponds to the code used to refer to the source in the catalog.

The goal of this paper is to update the ADHI catalog, with 500 h of observations at GBT (program GBT18A021) conducted in 2018 and 2019, with remeasured Nançay data from the Kinematics of the Local Universe (KLUN) and Nançay

Interstellar Baryons Legacy Extragalactic Survey (NIBLES) projects (Theureau et al. 1998, 2005, 2007, 2017; van Driel et al. 2008, 2016) retrieved from public repositories. This paper also presents rescaled newly available xArecibo Legacy Fast ALFA (ALFALFA) Arecibo data. (Haynes et al. 2018).

## 2. Measurement of the line width parameter

In the Cosmicflows program, we measured the neutral hydrogen line width at 21 cm wavelength enclosing 50% of the cumulative HI line flux,  $W_{m50}$  (Courtois et al. 2011). This parameter,  $W_{m50}$ , is the line width measured at the flux level that is 50% of the mean flux, averaged in channels within the wavelength range enclosing 90% of the total integrated flux. However, the parameter  $W_{m50}$  is only an empirical measure of the true width of an HI galaxy velocity profile. A correction for redshift and instrumental broadening should be applied:  $W_{m50}^c = \frac{W_{m50}}{1+z} - 2\Delta v\lambda$ , where  $z$  is the redshift,  $\Delta v$  is the smoothed spectral resolution, and  $\lambda = 0.25$  is an empirically determined constant. The observed line width can also be adjusted by separating out the broadening from turbulent motions and offsets to produce an approximation to  $2V_{\max}$ , where  $V_{\max}$  characterizes the rotation rate over the main body of a galaxy. Tully & Fouque (1985) defined the parameter  $W_{\text{mx}}$  as:

$$W_{\text{mx}}^2 = W_{m50}^2 + W_{t,m50}^2 \left[ 1 - 2e^{-(W_{m50}/W_{c,m50})^2} \right] - 2W_{m50}W_{t,m50} \left[ 1 - e^{-(W_{m50}/W_{c,m50})^2} \right]. \quad (1)$$

The parameters  $W_{c,m50} = 100 \text{ km s}^{-1}$  and  $W_{t,m50} = 9 \text{ km s}^{-1}$  were set after tests conducted in Courtois et al. (2009), and they

<sup>★</sup> Tables 2–4 are only available at the CDS via anonymous ftp to [cdsarc.u-strasbg.fr](https://cdsarc.u-strasbg.fr) (130.79.128.5) or via <http://cdsarc.u-strasbg.fr/viz-bin/cat/J/A+A/646/A113>

<sup>1</sup> [edd.ifa.hawaii.edu](http://edd.ifa.hawaii.edu)

**Table 1.** All digital HI catalog sources.

Code	Literature source
ksk2004	Koribalski et al. (2004)
shg2005	Springob et al. (2005)
hkk2005	Huchtmeier et al. (2005)
tmc2006	Theureau et al. (2006)
ghk2007	Giovanelli et al. (2007)
sgh2008	Saintonge et al. (2008)
kgh2008	Kent et al. (2008)
ctf2009	Courtois et al. (2009)
ctm2010-ctk2010	Courtois et al. (2011)
hgm2011	Haynes et al. (2011)
archi12	Courtois et al. (2015)
hsm2013	Hong et al. (2013)
tch2016	Theureau et al. (2017)
vbs2016	van Driel et al. (2016)
cdg2018	This paper

characterized the transition from boxcar to Gaussian intrinsic profiles and the turbulent broadening for the observed line width considered, respectively. It was then related to the rotation rate  $V_{\max}$  by:

$$2V_{\max} = \frac{W_{\max}}{\sin(i)}, \quad (2)$$

where  $i$  is the inclination of the galaxy from face-on relative to the observer. Inclinations have been evaluated using an online graphical tool, Galaxy Inclination Zoo (GIZ)<sup>2</sup>, in a collaborative science project with citizens. Please refer to Sect. 2.3 of Kourkchi et al. (2020b) for further details. Details regarding the  $W_{m50}$  and  $W_{\max}$  line width parameters and comparisons with alternatives are discussed in Courtois et al. (2009).

Finally, the error of the line width,  $e_w$ , is given by (Courtois et al. 2009):

$$\begin{aligned} S/N \geq 17 & \quad e_w = 8; \\ 2 < S/N < 17 & \quad e_w = 21.6 - 0.8 S/N; \\ S/N \leq 2 & \quad e_w = 70 - 25 S/N, \end{aligned} \quad (3)$$

where  $S/N$  is the signal-to-noise ratio ( $S/N$ ). An HI target is considered adequate for estimating its distance with the Tully-Fisher (TF).

Although the KLUN, NIBLES, and ALFALFA data have already been processed by their own collaborations, with methodologies described in their respective literature, these data samples have been reprocessed with the analysis described in this section in order to include them in the ADHI catalog, which gathers various HI data treated with that same procedure.

### 3. Data

#### 3.1. Nançay: NIBLES<sub>I</sub> and KLUN17

In this section, we consider two HI data releases that were obtained at the Nançay radio telescope by the KLUN and NIBLES collaborations: the KLUN17 release (Theureau et al. 2017) and the Nançay HI survey NIBLES<sub>I</sub> (van Driel et al. 2016). The previous releases of KLUN data have been already

<sup>2</sup> <http://edd.ifa.hawaii.edu/inclination/index.php>

included in the Cosmicflows catalogs, listed under the source code `tmc2006` in the ADHI catalog (see Table 1).

Figure 1 shows the sky distribution in equatorial coordinates of all HI data considered in this paper. The Nançay data considered in this section (both KLUN17 and NIBLES<sub>I</sub>) are represented by red squares. The other samples displayed in this figure are described further below in the next sections. The data points in the Nançay samples are distributed homogeneously in the two regions.

The redshift distribution of the two Nançay samples is shown in red in Fig. 2. Similarly to Fig. 1, the other HI samples are detailed later in the paper. We mainly find nearby galaxies, essentially below 6000 km s<sup>-1</sup>. The inset plot in the top right corner of Fig. 2 shows the redshift distribution of each Nançay subset. The purple and orange histograms represent the NIBLES<sub>I</sub> and KLUN17 subsamples, respectively. The NIBLES<sub>I</sub> sample primarily contains nearby galaxies below 3000 km s<sup>-1</sup>, and very few distant galaxies up to 8000 km s<sup>-1</sup>. In contrast, the KLUN17 data slightly contribute to the addition of distant galaxies for the next Cosmicflows catalog, as it mostly contains galaxies above 5000 km s<sup>-1</sup>, and up to 12 000 km s<sup>-1</sup>.

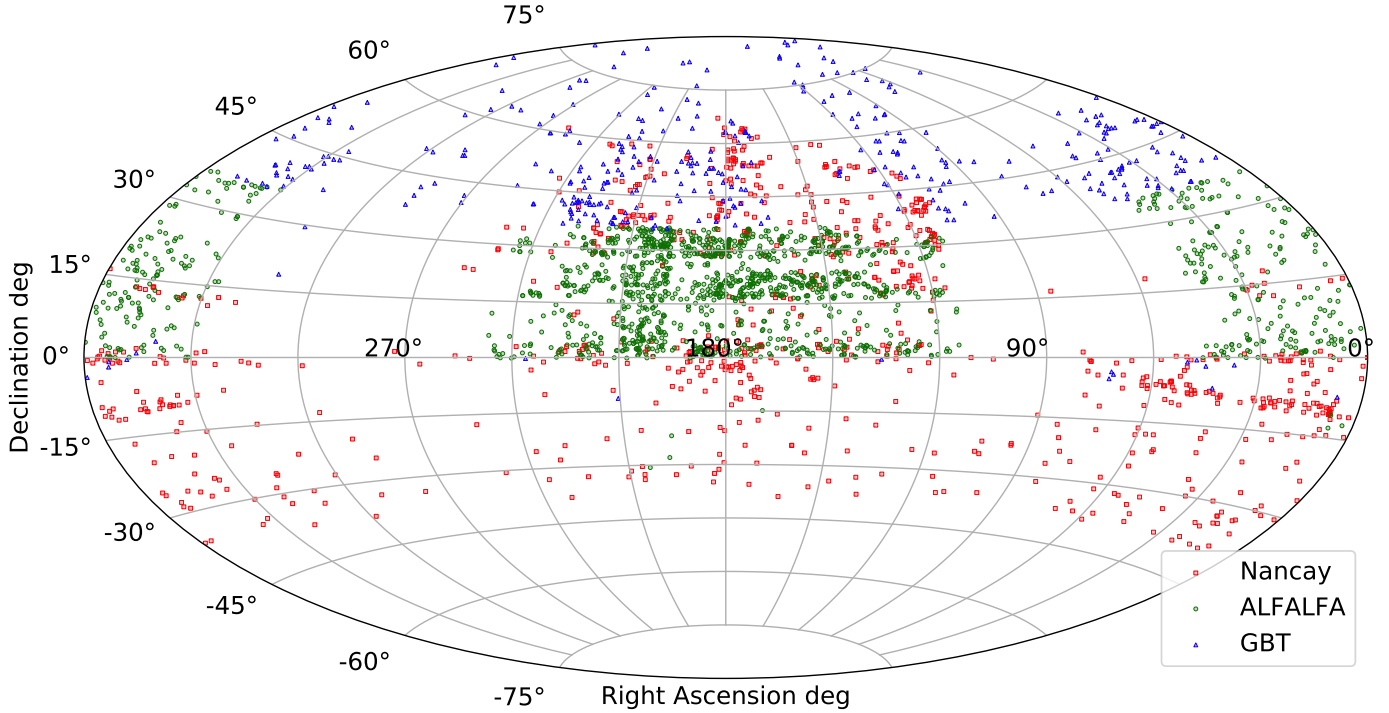
All data have been duly reduced and analyzed by their respective collaborations. However, for consistency, in order to include the line width measurements in the ADHI database, the HI parameter needs to be remeasured with the method used in the Cosmicflows collaboration. We reprocessed 1864 HI spectra from the NIBLES<sub>I</sub> collaboration. A total of 1443 line widths were detected and measured, including 808 new additions to the ADHI database, of which 565 are of an adequate quality for TF measurement. Regarding the KLUN17 sample, we obtained 500 detections after remeasuring the line width on 828 HI spectra. This provides 393 new additions to ADHI, of which 324 are of a sufficient high quality to be used for TF purposes.

The left panel of Fig. 3 shows the logarithm of the integrated flux as a function of the logarithm of the line width, and the right panel shows the distribution of HI mass as a function of the redshift. Identically to the previous Figs. 1 and 2, the Nançay data are represented by red squares, and the rest of the figures are described later in the paper.

As discussed above, line widths were remeasured with the original Nançay spectra in order to be integrated into the ADHI catalog. In Fig. 4, we compare the measurements published in this paper and added to ADHI as  $W^{\text{CF}}$  with the ones published by the NIBLES<sub>I</sub> and KLUN17 collaborations. These are indicated by  $W^{\text{published}}$  and represented as purple triangles and orange squares, respectively. The left panel shows the raw line widths  $W_{50}$ , while the right panel shows the line widths  $W_{\max}$  corrected with Eq. (1). In both panels, the remeasured  $W_{50}$  in the NIBLES<sub>I</sub> sample exhibits a significant positive offset of 11 km s<sup>-1</sup> from the originally published  $W_{50}$ . This is explained by the difference in measuring the width at 50% of the maximum flux and at 50% of the mean integrated flux. The KLUN sample shows consistency with the Cosmicflows measurements for both raw and corrected line widths.

#### 3.2. Arecibo: ALFALFA

The Arecibo Legacy Fast ALFA (ALFALFA) survey is a blind extragalactic HI survey to conduct a census of the local HI universe over a cosmologically significant volume. The full 100% ALFALFA extragalactic HI source catalog was published in Haynes et al. (2018). It contains more than 31 500 extragalactic HI line sources detected out to  $z < 0.06$  in the Arecibo telescope's declination range of  $0 < \delta < 38$  degrees.



**Fig. 1.** Sky distribution in equatorial coordinates of all HI data considered in this paper. The red squares, green dots, and blue triangles correspond to the HI data from the Nançay (NIBLES<sub>7</sub> AND KLUN17 HI data releases), Arecibo (ALFALFA), and Greenbank telescopes, respectively. All of the galaxies shown are new additions to the ADHI catalog, and their line width measurements are adequate for the use of the TF relation.

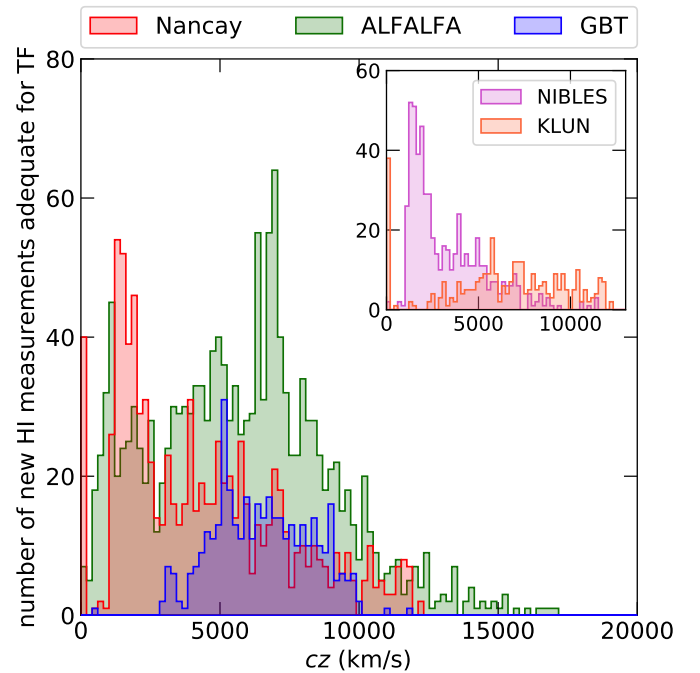
### 3.2.1. Identifying galaxies with adequate measurements

The various HI parameters obtained by the ALFALFA collaboration have been published and are available online for 32 612 galaxies. However, to date, only 40% of the actual HI spectra are available in a digital usable format (e.g., ascii, fits) ( $\alpha$ 40, 15 000 galaxies, published in Haynes et al. 2011). These data were already remeasured by the Cosmicflows collaboration, bringing a total of 3898 new good quality line width measurements to the ADHI catalog out of 15 000 reprocessed spectra (Courtois & Tully 2015). The source code hgm2011 listed in Table 1 corresponds to the  $\alpha$ 40 entry in the ADHI catalog. However, the remaining ALFALFA spectra of the full  $\alpha$ 100 catalog were not available in the  $\alpha$ 40 release, so the HI parameters cannot be directly re-estimated and included in ADHI.

In this section, we compare the parameters from the  $\alpha$ 40 catalog obtained by ALFALFA and Cosmicflows. After establishing relations between these parameters, we then estimate the values of the HI parameters of the whole ALFALFA-100% catalog, as if they were remeasured by the Cosmicflows collaborations. Throughout this section, parameters obtained by the ALFALFA (published in the  $\alpha$ 40 catalog) and Cosmicflows collaborations are identified by the subscripts  $\alpha$  and CF, respectively.

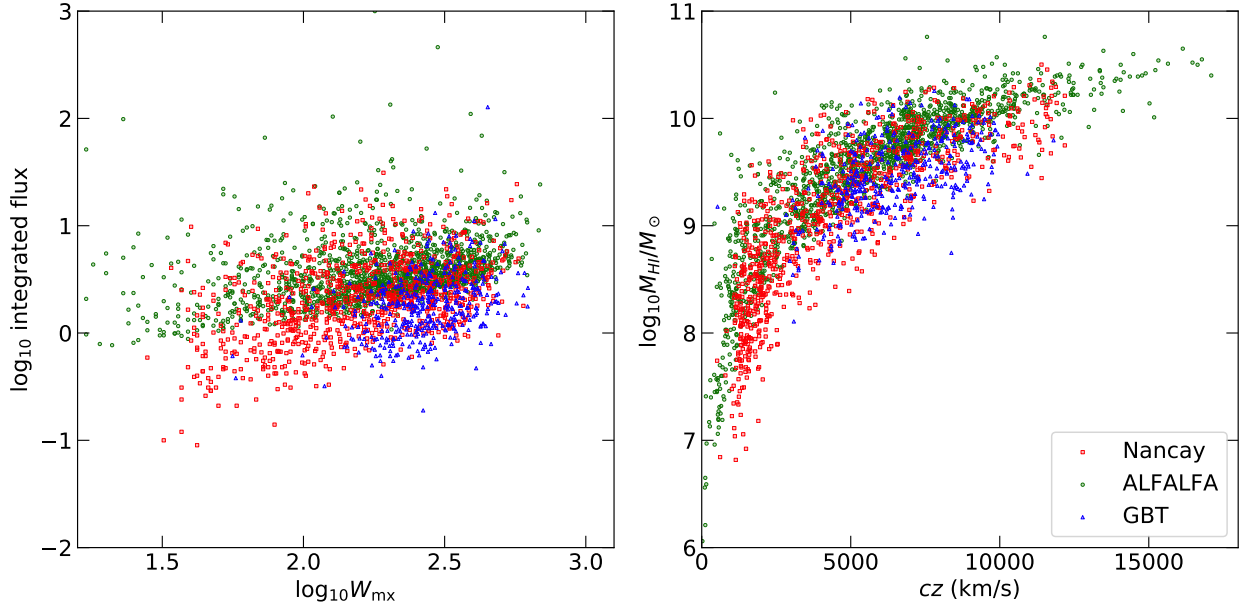
As our goal is to estimate which ALFALFA-100% line width measurements may be adequate for deriving distances, we first consider the error,  $e_W$ , on the HI line width. The definition of this parameter differs between the two studies. When comparing the two parameters  $e_W^\alpha$  and  $e_W^{CF}$ , one can easily notice that  $e_W^\alpha \neq e_W^{CF}$ .

The S/Ns,  $S/N^{CF}$ , and,  $S/N^\alpha$ , are defined differently. In the Cosmicflows collaboration, the S/N is derived as the ratio of the signal at 50% of the mean flux over the noise measured beyond the extremities of the signal (Courtois & Tully 2015). The S/N derived by the ALFALFA collaboration is defined in Eq. (2) of Haynes et al. (2011). Both S/Ns,  $S/N^{CF}$  and  $S/N^\alpha$ , are compared

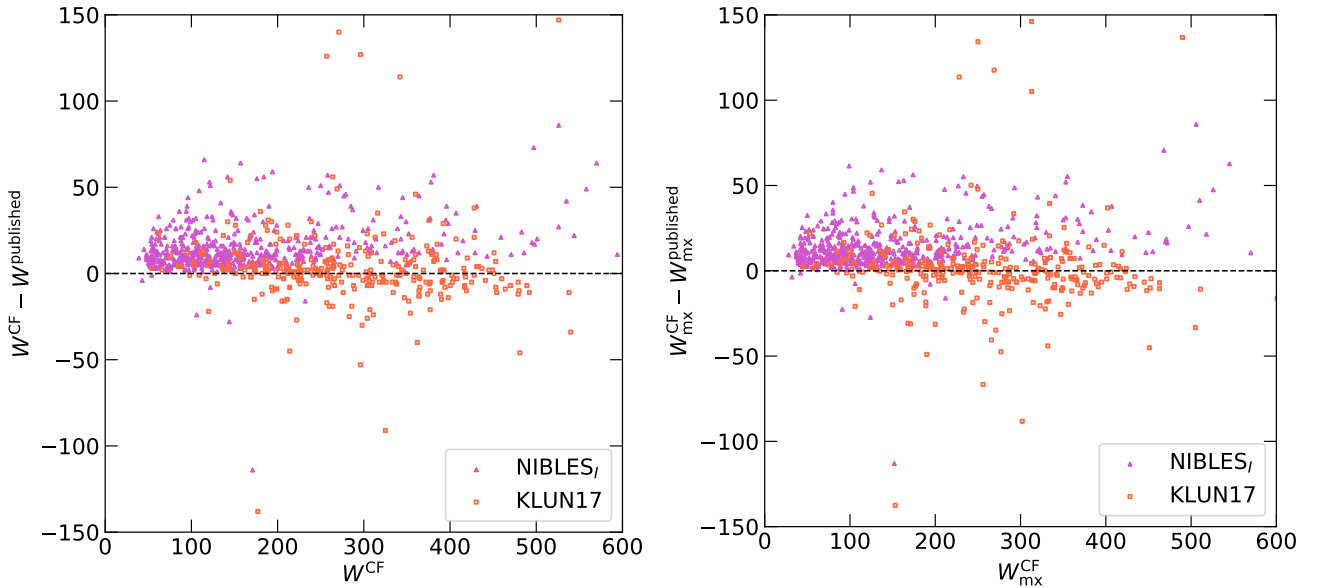


**Fig. 2.** Redshift distribution of new HI data described in this paper. The Nançay, Arecibo, and GBT data are shown in red, green, and blue, respectively. Only new additions to the ADHI catalog that are adequate for the use of the TF relation are considered. The subset plot in the top right corner shows the redshift distribution of Nançay data only. The NIBLES<sub>7</sub> and KLUN17 data are in purple and orange, respectively.

directly in the left panel of Fig. 5. Only galaxies adequate for TF, that is to say when  $e_W^{CF} \leq 20$ , are considered. The black dotted line represents the case when  $S/N^{CF} = S/N^\alpha$ . One can relate the



**Fig. 3.** *Left:* logarithm of the integrated flux as a function of the logarithm of the line width. *Right:* distribution of HI mass as a function of the redshift. Nancay data (NIBLES<sub>I</sub> and KLUN17) are represented by red squares, Arecibo data by green circles, and GBT data by blue triangles. Only new additions to the ADHI catalog that are adequate for the use of the TF relation are considered.



**Fig. 4.** Comparison of the measurements published in this paper and added to ADHI as  $W^{\text{CF}}$  to the ones published in the NIBLES<sub>I</sub> and KLUN17 data releases, represented by  $W^{\text{published}}$ . The data obtained by the NIBLES<sub>I</sub> and KLUN17 collaborations are represented by purple triangles and orange squares, respectively. The black dotted line represents the case when  $W^{\text{cf}} = W^{\text{published}}$ . *Left:* comparison of the raw line widths  $W_{50}^{\text{CF}}$  and  $W_{50}^{\text{published}}$ . *Right:* comparison of the corrected line widths  $W_{\text{mx}}^{\text{CF}}$  and  $W_{\text{mx}}^{\text{published}}$ . In both panels, we observe a significant offset of 11 km s<sup>-1</sup> for the NIBLES<sub>I</sub> sample.

S/Ns from both collaborations by the following relation:

$$S/N^\alpha = 1.7 + 2.9 S/N^{\text{CF}}, \quad (4)$$

which is illustrated in the left panel of Fig. 5 by a solid red line. We can then estimate the value of  $S/N^\alpha$  in the Cosmicflows definition, which is denoted by  $S/N^{\alpha \rightarrow \text{CF}}$  hereafter.

### 3.2.2. Comparing the ALFALFA and Cosmicflows line widths

The line widths obtained with the 70% completion ALFALFA release ( $\alpha 70$ ) and Cosmicflows catalogs have been compared

without reprocessing the spectra in Kourkchi et al. (2019), using a subsample of galaxies with adequate quality HI measurements. They were part of both catalogs at that time. The result was  $W_{\text{mx}}^{\text{CF}} = W_{50}^\alpha - 6$ , where  $W_{\text{mx}}^{\text{CF}}$  is the line width in km s<sup>-1</sup> taken from ADHI (including corrections, see Eq. (1)), and  $W_{50}^\alpha$  is the line width provided by the ALFALFA  $\alpha 70$  catalog.

In this section, we follow the same methodology to compare line widths from both collaborations, considering now the newly published full 100% ALFALFA catalog Haynes et al. (2018) which contains 32 612 galaxies. Line widths extracted from this catalog are noted as  $W_{50}^\alpha$ . We are comparing them to a sample

containing 15 433 good quality only line width measurements extracted from the ADHI dataset, corrected with Eq. (1) and represented by  $W_{\text{mx}}^{\text{CF}}$ . The number of common galaxies in both datasets is 3970.

In the right panel of Fig. 5, we compare the line widths  $W_{50}^{\alpha}$  and  $W_{\text{mx}}^{\text{CF}}$  by plotting the difference  $W_{\text{mx}}^{\text{CF}} - W_{50}^{\alpha}$  as a function of the S/N,  $S/N^{\alpha}$ . We notice that there is significant dispersion at small S/Ns, where  $S/N^{\alpha} < 15$ , which is represented by a black dotted line. Line width measurements above this threshold yield the following linear relationship between the two parameters:

$$W_{\text{mx}}^{\text{CF}} = W_{50}^{\alpha} - 2 \text{ (km s}^{-1}\text{)} \quad (5)$$

represented in the right panel by a solid red line. This relation is slightly different from the one obtained previously in Kourkchi et al. (2019). We attribute this slight difference to the use of an earlier data release of the ALFALFA survey,  $\alpha 70$ , by Kourkchi et al. (2019).

### 3.2.3. Application to the full ALFALFA data

In this section, Eq. (5) is applied to the full 100% ALFALFA catalog (galaxies previously included in ADHI with high quality measurements are excluded) to compute how many and which galaxies could be added to the upcoming Cosmicflows-4 catalog. Galaxies with HI line width errors (in the Cosmicflows definition) of  $e_W^{\alpha \rightarrow \text{CF}} \leq 20 \text{ km s}^{-1}$  are suitable for the TF method and as such for inclusion in the new upcoming Cosmicflows-4 catalog. We identify 1515 new galaxies which satisfy this criterion.

Figure 1 displays the sky distribution of the additional 1515 ALFALFA galaxies that are involved in the construction of the upcoming CF4 catalog. ALFALFA galaxies are represented by green dots. A homogeneous coverage can be easily noticed in the region covered by Arecibo.

Figure 2 shows the redshift distribution of the new galaxies from ALFALFA in green. Three peaks are observed at  $2000 \text{ km s}^{-1}$ ,  $5000 \text{ km s}^{-1}$ , and  $8000 \text{ km s}^{-1}$ . This sample will thus be a significant contribution to CF4. It will include a large number of distant galaxies up to  $17\,000 \text{ km s}^{-1}$ . However, it is important to note that only 338 galaxies ( $\sim 22\%$  of the sample) have a redshift larger than  $8000 \text{ km s}^{-1}$ .

The left panel of Fig. 3 shows the integrated flux as a function of the line width. ALFALFA galaxies are represented by green dots. The blind ALFALFA survey mostly detects galaxies near the peak of the HI luminosity function where the largest volume is explored at a given flux. Hence, detection is favored for giant galaxies with a large line width and intrinsic HI fluxes. Smaller galaxies at the same apparent flux levels are drawn from much smaller volumes.

The distribution of the HI mass as a function of the redshift is shown in the right panel of Fig. 3. ALFALFA galaxies are once again displayed as green dots. The luminosity function (LF) of galaxies is defined as the number of galaxies per  $\text{Mpc}^3$  in a luminosity interval  $dM$  centered on magnitude  $M$  (Zwaan et al. 2001). In Fig. 3, we observe that the new galaxies observed satisfactorily with GBT are significantly less massive than the mean (or typical) ones from ALFALFA. The ALFALFA remeasurements are consistent in terms of the luminosity function with other surveys. The point concerning our GBT observations is that we do not aim to observe a representative survey of galaxies showing a good sample of a different luminosity or mass distribution galaxies. Our goal is to measure galaxies as far as possible, regarding their velocity recession, in order to obtain an enlarged distribution of velocities with independent

distances (using the TF extraction), for our further studies of peculiar velocity. In other terms, we are not looking for a representative sample of galaxies following a luminosity function, rather we target galaxies with spatial, recession velocity and distances as various as possible.

## 3.3. Green Bank telescope

### 3.3.1. Selection of targets

Targets were selected from a compilation of two samples: a sample of flat galaxies from the Revised Flat Galaxy Catalog (RFGC Karachentsev et al. 1999) to which a selection cut of  $\delta > 36$  deg was applied to avoid overlap with the ALFALFA sample; and a sample of galaxies near to the Dipole Repeller (Hoffman et al. 2017, DR) – galaxies were extracted from the LEDA database<sup>3</sup> (Makarov et al. 2014) with the cuts  $\alpha > 20$  deg,  $16 < \delta < 65$  deg,  $3,000 < cz < 12\,000 \text{ km s}^{-1}$ , and  $i > 50$  deg.

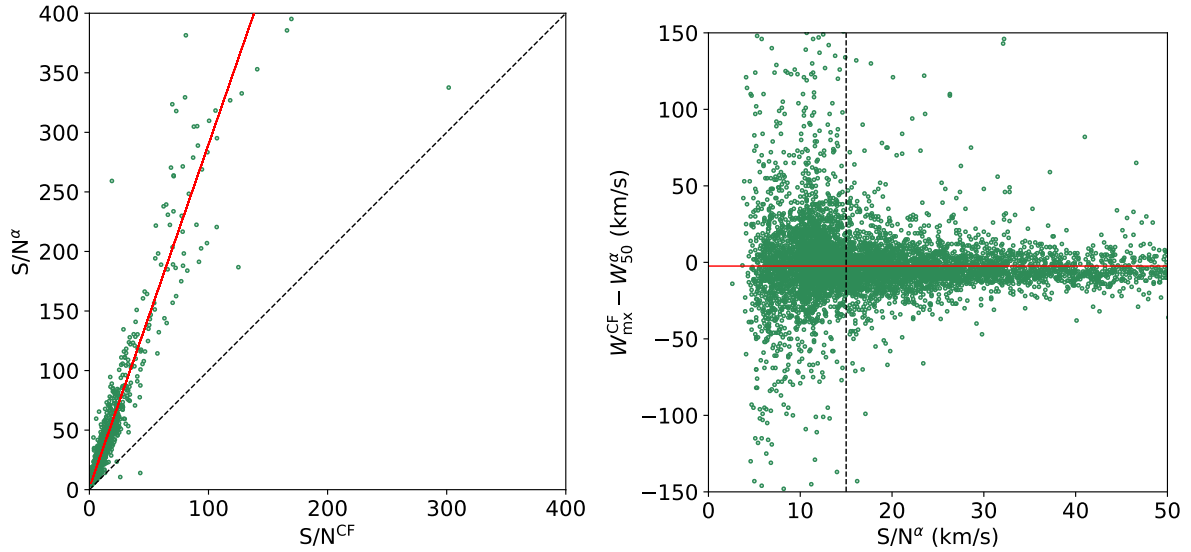
Several samples were added later to augment the number of potential targets. First, a sample of spiral galaxies extracted from LEDA with the following cuts applied:  $\delta > 36$  deg (no overlap with ALFALFA),  $3000 < cz < 9000 \text{ km s}^{-1}$ , and  $i > 55$  deg. The cuts were later modified in order to obtain more distant galaxies:  $\delta > 38$  deg,  $3000 < cz < 10\,000 \text{ km s}^{-1}$ , and  $40 < i < 85$  deg. Lastly, a few SNIa hosts from SNfactory (Rigault et al. 2020; Aldering et al. 2002) with a redshift up to  $12\,000 \text{ km s}^{-1}$  were added to the list of targets.

All galaxies that are already part of the ADHI catalog with adequate HI profiles were removed from the samples described above. Several galaxies with almost adequate HI profiles from the GB300 and Nançay telescopes have also been added to the list of potential targets.

This extensive list now contains 5460 galaxies, all of which have been inspected and are observed if they satisfy the following criteria. First, we needed to make sure that the galaxies were spiral galaxies with HI gas. One can look at the Pan-STARRS (Chambers et al. 2016) optical images in the  $g, r, i, z$ , and  $y$  bands. A galaxy with HI, especially its disk, is hardly visible in the  $y$  band, while it is very bright at the  $g$  band. We also searched the optical spectra from SDSS DR12 (Alam et al. 2015) or LAMOST (Luo et al. 2015) for a significant H $\alpha$  emission line, hinting at the presence of young stars and therefore HI gas. Spectra containing Na, Mg, and Ca emission lines are indicative of old stellar populations and little HI. Available Pan-STARRS images and SDSS or LAMOST spectra were systematically inspected by eye for each target. A total of 54% of the targets had an optical spectrum available. This ratio is not entirely satisfactory, but we did not find a recent, available optical spectrum for all our targets. Secondly, we did not observe collisions or interacting galaxies. If the distribution of the HI gas is disturbed, it does not allow us to measure the rotation speed of each galaxy. Moreover, the size of the radio lobe must be considered. Two galaxies within the radio lobe and located at the same redshift would lead to a confused spectrum because the galaxies spectra would overlap. Last but not least, we also checked the photometry quality and the inclination of the potential target to ensure the TF relation is applicable. Inclinations are indispensable to de-project the HI line widths. We require that inclinations be greater than 45 degrees from face-on.

After inspection of 5460 potential targets, 628 galaxies were selected and observed. The target selection procedure was not completely carried out in the sense that we observed

<sup>3</sup> <http://leda.univ-lyon1.fr/>



**Fig. 5.** Comparison of HI parameters derived by the Cosmicflows and ALFALFA collaborations identified by the subscripts CF and  $\alpha$ , respectively. Only galaxies available in the  $\alpha 40$  release are considered. *Left:* comparison of the S/Ns,  $S/N^{CF}$  and  $S/N^\alpha$ . *Right:* comparison of the line widths  $W_{mx}^{CF}$  and  $W_{50}^\alpha$ . The difference  $W_{mx}^{CF} - W_{50}^\alpha$  is represented as a function of the S/N,  $S/N^\alpha$ . The black dotted line represents the threshold  $S/N^\alpha = 15$  discussed in the text. Data above this threshold can be fitted to  $W_{mx}^{CF} = W_{50}^\alpha - 2 \text{ km s}^{-1}$ , which is represented by a solid red line.

**Table 2.** Raw HI parameters measured on all 628 spectra obtained at GBT as well as complementary data.

PGC	Name	$V_{hel}$	$V_{50}$	$W_{50}$	$e_w$	$S/N$	Flux	$i$	$b_t$
676	UGC 00085	5153	5157	277	16	6.9	3.7	48	14.46
22698	UGC 04205	6850	6844	390	18	4.1	3.1	90	15.22
20121	UGC 03655	6170	6056	201	49	0.8	1.5	42	14.49
9432	IC 1799	5022	4926	422	100	6.9	6.6	64	14.89
8177	UGC 01628	5556	5650	76	500	0.9	0.3	62	15.65

**Notes.** The full table is available at the CDS. The error  $e_w = 100 \text{ km s}^{-1}$  corresponds to the confusion case where, for example, two galaxies are in the detection field,  $e_w = 500 \text{ km s}^{-1}$ , which corresponds to the no detection case.

well-defined targets with our first selection of criteria but which ultimately do not give a satisfactory result after observation.

### 3.3.2. Observations: Strategy and planning

A total of 610h of observations have been conducted from December 2017 to May 2019. During observations, the telescope was configured as follows. We used the  $L$ -band receiver Rcvr1\_2, detecting frequencies between 1.15 GHz and 1.73 GHz. For this receiver, the gain of the GBT is  $2 \text{ K Jy}^{-1}$ , and the size of the radiolobe is 9 arcmin. The spectral resolution considered during observations is  $0.9 \text{ km s}^{-1}$ .

The following ON-OFF-ON methodology was used throughout our observations. This observing strategy allows two ON-OFF scans to be conducted on the same source during a single session of observation. It consists of one 300 s scan on the source (ten integrations of 30 s), followed by a 300 s scan on the sky, then a second 300 s scan on the source. A total of 15 min is spent on a single source during a session. This strategy permits an additional 5 min of observation time per source and per session compared to the usual ON-OFF methodology.

During an observation run, targets are selected automatically from the list of inspected targets based on their proximity relative to the telescope orientation. Other target selection criteria are  $0 < \delta < 85 \text{ deg}$  and that the angular separation between the target and the Sun should be higher than 10 deg. After each run, we

reduced the collected data as described in Sect. 3.3.3 below and updated the target list, accordingly. Keywords are used to label and identify galaxies to be observed. The ON-OFF-ON cycles are repeated until the signal is sufficient to give an adequate line width measurement ( $S/N$  high enough to give an acceptable error on the line width, see Sect. 2). The total integration time per source goes from 15 min (one ON-OFF-ON cycle), to several hours for the most distant galaxies, or when the spectra were contaminated by radio frequency interferences (RFIs) as detailed in Sect. 3.3.3 below.

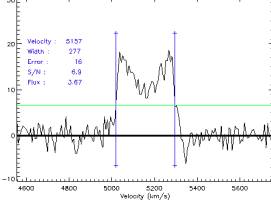

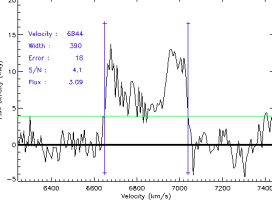
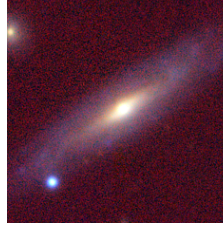
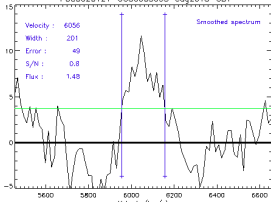
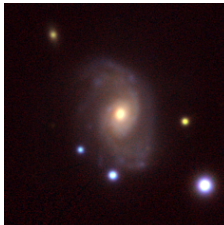
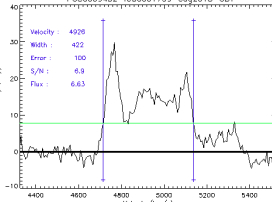
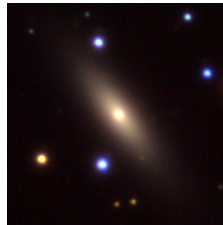
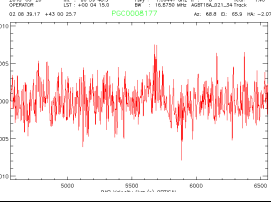
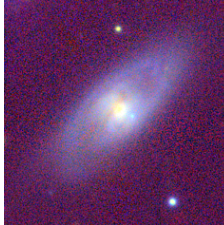
### 3.3.3. Data reduction and RFIs

Each integration of a single ON-OFF pair is calibrated with the `getsigref` function of GBTIDL<sup>4</sup> in order to obtain the final spectra. After removing eventual RFIs as described in this section below, a Hanning smoothing (hanning function) is applied in order to obtain a resolution of  $3.6 \text{ km s}^{-1}$ .

In total, 628 HI spectra of galaxies have been obtained at GBT, of which 407 correspond to detections and 385 are acceptable for the use of the TF relation. The raw HI parameters measured on the 628 spectra obtained are available in Table 2, as well as complementary data. Column 1 corresponds to the ID number of the target in the Principal Galaxy catalog (PGC).

<sup>4</sup> <http://gbtidl.nrao.edu>

**Table 3.** HI profiles and PanSTARRS multi-band images of the 628 galaxies observed at GBT.

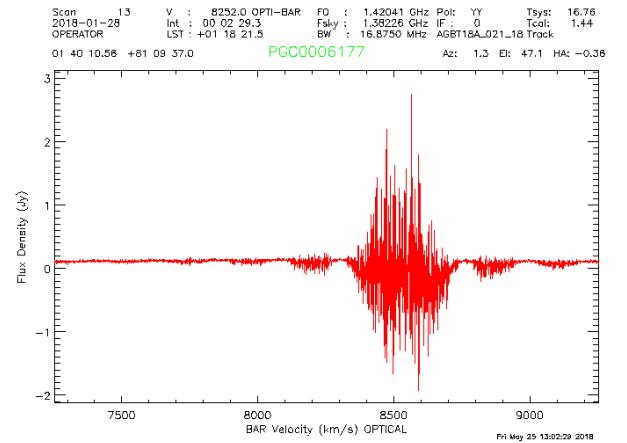
PGC/ $V_{\text{hel}}$ Profile	$W_{50}/e_w$ PanSTARRS	PGC/ $V_{\text{hel}}$ Profile	$W_{50}/e_w$ PanSTARRS
PGC0000676 / 5157 	277 / 16 	PGC0022698 / 6844 	390 / 18 
PGC0020121 / 6056 	201 / 49 	PGC0009432 / 4926 	422 / 100 
PGC0008177 / 5650 	76 / 500 		

**Notes.** Columns 1 and 3 give the PGC number, the heliocentric velocity  $V_{\text{hel}}$  in  $\text{km s}^{-1}$ , and the HI profile after the measurement of the line width. In the case of non-detection, the raw spectra obtained from the GBT is shown in red. Columns 2 and 4 give the raw line width  $W_{50}$  and its error  $e_w$  in  $\text{km s}^{-1}$ , as well as the PanSTARRS image composite of the bands  $y/i/g$ . The full table is available at the CDS.

Column 2 gives an alternative name. Columns 3 and 4 correspond to the input (heliocentric) velocity  $V_{\text{hel}}$  and the velocity measured on the profile  $V_{50}$ . The line width  $W_{50}$  and its uncertainty  $e_w$ , as well as the S/N and the measured HI line flux are provided in Cols. 5–8, respectively. We note that we use the error codes  $e_w = 100 \text{ km s}^{-1}$  and  $e_w = 500 \text{ km s}^{-1}$  for confused spectra and non-detections, respectively. The inclination is listed in Col. 9, and the total  $B$ -band magnitude is provided in Col. 10. All non-HI parameters (input velocity, inclination, and magnitude) have been extracted from the LEDA database. All HI profiles and corresponding PanSTARRS optical images of the 628 galaxies observed can be found in Table 3, which is available electronically at the journal’s homepage.

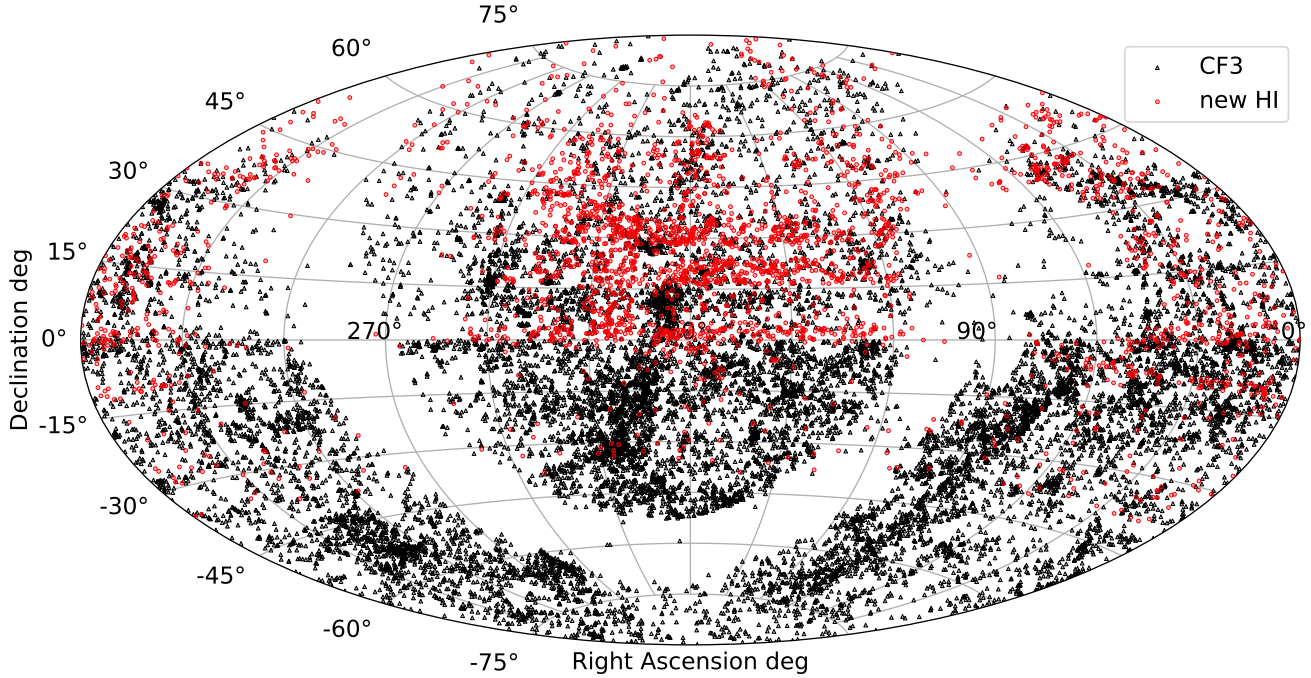
The main problems faced during data reduction arose from RFI. The single peak RFIs are removed easily by interpolation and do not affect the final spectra and measurements of HI parameters.

However, a significant RFI located at  $8500 \text{ km s}^{-1}$  (or  $1.381 \text{ MHz}$ ) has been encountered frequently. It may be generated by a radar or a GPS system used for nuclear tests. This RFI destroyed a significant amount of HI spectra of distant galaxies located between  $8000 \text{ km s}^{-1}$  and  $9000 \text{ km s}^{-1}$ . The RFI is visible, for example, on the HI spectra of PGC6177 located at  $cz = 8252 \text{ km s}^{-1}$ , as shown in Fig. 6. The HI line on contaminated spectra is not visible at all as the amplitude of the RFI (up to  $\approx 10 \text{ Jy}$ ) is much larger than the usual amplitude of an HI line



**Fig. 6.** HI spectra of PGC6177 located at  $cz = 8,252 \text{ km s}^{-1}$ . A strong radio frequency interference (RFI) occurring frequently at  $8500 \text{ km s}^{-1}$  can be noticed, while the HI line is not visible.

( $\approx 15 \text{ mJy}$ ). We chose to ignore integrations of a single scan contaminated by this RFI when deriving the final HI spectra. The FFT methodology (sigma clipping in Fourier space) suggested by Hunt et al. (2016) has been tested on contaminated spectra with no success. Please refer to this publication for more details on the method.



**Fig. 7.** Sky distribution in equatorial coordinates of all new line width measurements presented in this paper, which are adequate for TF distance measurement. The black triangles correspond to the Cosmicflows-3 catalog, while the red dots represent the new additional HI data.

Finally, 192 targets have been observed for which no HI lines have been detected, and 29 spectra are confused. Sadly, a significant amount of observing time has been lost on these targets, even though all targets have been deeply inspected prior to observing to check if an HI disk may be present.

### 3.4. New additions to the ADHI collection

The sky distribution of the 385 new additions to the ADHI collection is shown in Fig. 1. GBT galaxies are displayed as blue triangles. We focused on the northern celestial hemisphere to address the lack of the data coverage by Cosmicflows-3 in this region.

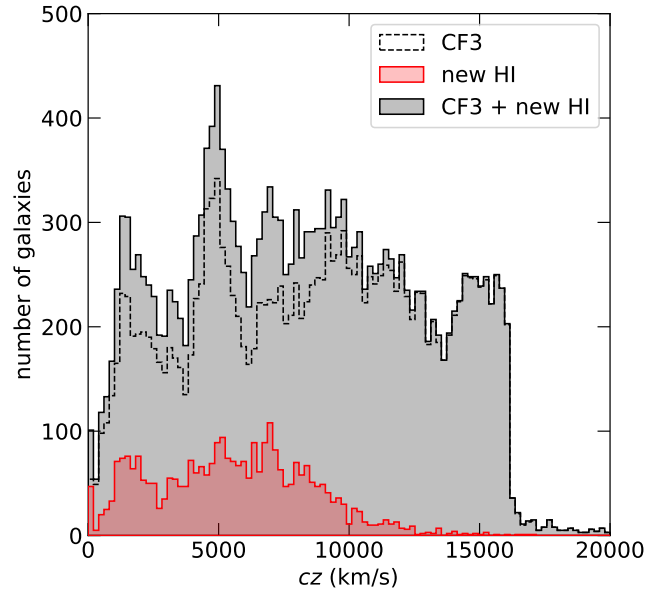
The redshift distribution of these galaxies is represented in blue in Fig. 2. Mostly distant galaxies (up to  $10\,000\text{ km s}^{-1}$ ) have been observed and added to ADHI.

The left panel of Fig. 3 shows the log of the integrated flux as a function of the log of the HI line width. We mostly observed massive galaxies with low flux. The right panel shows the distribution of the HI mass as a function of the redshift.

## 4. From Cosmicflows-3 to Cosmicflows-4

The sky distribution of all new line width measurements adequate for TF is shown in Fig. 7. The black triangles correspond to the CF3 catalog, while the red dots represent the new HI data. Most of the new data are located in the north in order to correct the north-south asymmetry in Cosmicflows-3. The imbalance in the number of galaxies is indeed slightly corrected. In CF3,  $\sim 70\%$  of the data were located in the south and only  $\sim 30\%$  in the north. After adding the new HI data presented in this paper,  $\sim 55\%$  of data should be in the south and  $\sim 45\%$  in the north.

In Fig. 8, we compare the redshift distributions of the CF3 data and the new HI data to be added to CF4. The CF3 data are represented in gray with a black dotted line, while the new HI data are over-plotted and shown in red. The combination of CF3



**Fig. 8.** Redshift distribution of CF3 data (black dotted line) and resulting addition into CF4 catalog (merging of CF3 and new HI data, gray with solid black line). The redshift distribution of the new HI data is shown in red.

with these new data to produce CF4 is shown in Fig. 8 in gray with a solid black line. The contribution of the TF method to CF4 becomes less significant as  $cz$  increases. The new HI data added in the north are not distant enough to fully compensate for the lack of data in this region, compared to the south which is mainly covered by the fundamental plane method with 6dF.

We have updated the ADHI catalog by incorporating all galaxies with new HI observations at GBT and those with reprocessed Nançay HI measurements presented in this study, as listed in Table 4. Column 1 corresponds to the ID number of the target

**Table 4.** Newly observed and reprocessed HI measurements presented in this paper.

PGC	Name	Source	Telescope	$V_{\text{hel}}$	$W_{\text{mx}}$	$e_W$	$S/N$
14330	MCG-03-11-007	Theureau et al. (2017)	Nançay	7536	376	18	3.3
25180	MCG+03-23-021	Theureau et al. (2017)	Nançay	6150	194	17	5.7
1862	MCG-02-02-033	van Driel et al. (2016)	Nançay	5351	266	18	4.2
43634	UGC08012	This paper	GBT	8160	290	18	3.5
47833	MCG+06-30-042	This paper	GBT	7005	196	17	5.4

**Notes.** The full table is available at the CDS.

in the Principal Galaxy catalog (PGC). An alternative name is given in Col. 2. The source of the spectra and the telescope used to conduct the observation are given in Cols. 3 and 4, respectively. The heliocentric velocity  $V_{\text{hel}}$  is provided in Col. 5. Main HI parameters are listed in Cols. 6–8.

## 5. Conclusion

We have acquired 385 high quality HI measurements from 628 targets using the GBT and 889 good measurements of the HI line width from 2692 remeasured Nançay spectra. These 1274 newly observed and reprocessed galaxies have been added to our ADHI catalog maintained at the Extragalactic Distance Database (EDD). Furthermore, we identify 1515 new ALFALFA galaxies with spectra that may be sufficient for TF distance measurement. In total, this paper brings an additional 2789 new galaxies that were not previously in our All Digital Catalog.

However, this study does not fully counterbalance the celestial north-south asymmetrical coverage of the HI observations in the Cosmicflows catalogs. In the north, future HI surveys, such as Apertif in the Netherlands and FAST in China, could potentially improve this imbalance.

*Acknowledgements.* HC is grateful to the Institut Universitaire de France and CNES for its support. AD acknowledges financial support from the Project IDEXLYON at the University of Lyon under the Investments for the Future Program (ANR-16-IDEX-0005). We acknowledge the use of the HyperLeda database. This project is partly financially supported by Region Rhone-Alpes-Auvergne.

## References

Alam, S., Albareti, F. D., Allende Prieto, C., et al. 2015, *ApJS*, **219**, 12  
 Aldering, G., Adam, G., Antilogus, P., et al. 2002, in *Overview of the Nearby Supernova Factory*, eds. J. A. Tyson, & S. Wolff, *SPIE Conf. Ser.*, **4836**, 61  
 Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, ArXiv e-prints [arXiv:1612.05560]  
 Courtois, H. M., & Tully, R. B. 2015, *MNRAS*, **447**, 1531

Courtois, H. M., Tully, R. B., Fisher, J. R., et al. 2009, *AJ*, **138**, 1938  
 Courtois, H. M., Tully, R. B., Makarov, D. I., et al. 2011, *MNRAS*, **414**, 2005  
 Courtois, H. M., Zaritsky, D., Sorce, J. G., & Pomarède, D. 2015, *MNRAS*, **448**, 1767  
 Courtois, H. M., Tully, R. B., Hoffman, Y., et al. 2017, *ApJ*, **847**, L6  
 Giovanelli, R., Haynes, M. P., Kent, B. R., et al. 2007, *AJ*, **133**, 2569  
 Haynes, M. P., Giovanelli, R., Martin, A. M., et al. 2011, *AJ*, **142**, 170  
 Haynes, M. P., Giovanelli, R., Kent, B. R., et al. 2018, *ApJ*, **861**, 49  
 Hoffman, Y., Pomarède, D., Tully, R. B., & Courtois, H. M. 2017, *Nat. Astron.*, **1**, 0036  
 Hong, T., Staveley-Smith, L., Masters, K. L., et al. 2013, *MNRAS*, **432**, 1178  
 Huchtmeier, W. K., Karachentsev, I. D., Karachentseva, V. E., Kudrya, Y. N., & Mitronova, S. N. 2005, *A&A*, **435**, 459  
 Hunt, L. R., Pisano, D. J., & Edell, S. 2016, *AJ*, **152**, 30  
 Karachentsev, I. D., Karachentseva, V. E., Kudrya, Y. N., Sharina, M. E., & Parnovskij, S. L. 1999, *Bull. Spec. Astrophys. Obs.*, **47**, 5  
 Kent, B. R., Giovanelli, R., Haynes, M. P., et al. 2008, *AJ*, **136**, 713  
 Koribalski, B. S., Staveley-Smith, L., Kilborn, V. A., et al. 2004, *AJ*, **128**, 16  
 Kourkchi, E., Tully, R. B., Neill, J. D., et al. 2019, *ApJ*, **884**, 82  
 Kourkchi, E., Courtois, H. M., Graziani, R., et al. 2020a, *AJ*, **159**, 67  
 Kourkchi, E., Tully, R. B., Eftekharzadeh, S., et al. 2020b, *ApJ*, **902**, 145  
 Luo, A.-L., Zhao, Y.-H., Zhao, G., et al. 2015, *Res. Astron. Astrophys.*, **15**, 1095  
 Makarov, D., Prugniel, P., Terekhova, N., Courtois, H., & Vauglin, I. 2014, *A&A*, **570**, A13  
 Rigault, M., Brinnet, V., Aldering, G., et al. 2020, *A&A*, **644**, A176  
 Saintonge, A., Giovanelli, R., Haynes, M. P., et al. 2008, *AJ*, **135**, 588  
 Springob, C. M., Haynes, M. P., Giovanelli, R., & Kent, B. R. 2005, *ApJS*, **160**, 149  
 Theureau, G., Bottinelli, L., Coudreau-Durand, N., et al. 1998, *A&AS*, **130**, 333  
 Theureau, G., Coudreau, N., Hallet, N., et al. 2005, *A&A*, **430**, 373  
 Theureau, G., Martin, J. M., Cognard, I., & Borsenberger, J. 2006, in *Astronomical Data Analysis Software and Systems XV*, eds. C. Gabriel, C. Arviset, D. Ponz, & S. Enrique, *ASP Conf. Ser.*, **351**, 429  
 Theureau, G., Hanski, M. O., Coudreau, N., Hallet, N., & Martin, J. M. 2007, *A&A*, **465**, 71  
 Theureau, G., Coudreau, N., Hallet, N., Hanski, M. O., & Poulain, M. 2017, *A&A*, **599**, A104  
 Tully, R. B., & Fouque, P. 1985, *ApJS*, **58**, 67  
 Tully, R. B., Courtois, H., Hoffman, Y., & Pomarède, D. 2014, *Nature*, **513**, 71  
 Tully, R. B., Pomarède, D., Graziani, R., et al. 2019, *ApJ*, **880**, 24  
 van Driel, W., Schneider, S. E., Lehnert, M., & Minchin, R. 2008, in *The Evolution of Galaxies Through the Neutral Hydrogen Window*, eds. R. Minchin, & E. Momjian, *Am. Inst. Phys. Conf. Ser.*, **1035**, 256  
 van Driel, W., Butcher, Z., Schneider, S., et al. 2016, *A&A*, **595**, A118  
 Zwaan, M. A., Briggs, F. H., & Sprayberry, D. 2001, *MNRAS*, **327**, 1249