UNIVERSITÉ PARIS SUD
MAGISTÈRE DE PHYSIQUE FONDAMENTALE
BÂT. 470 - 91405 ORSAY CEDEX
FRANCE

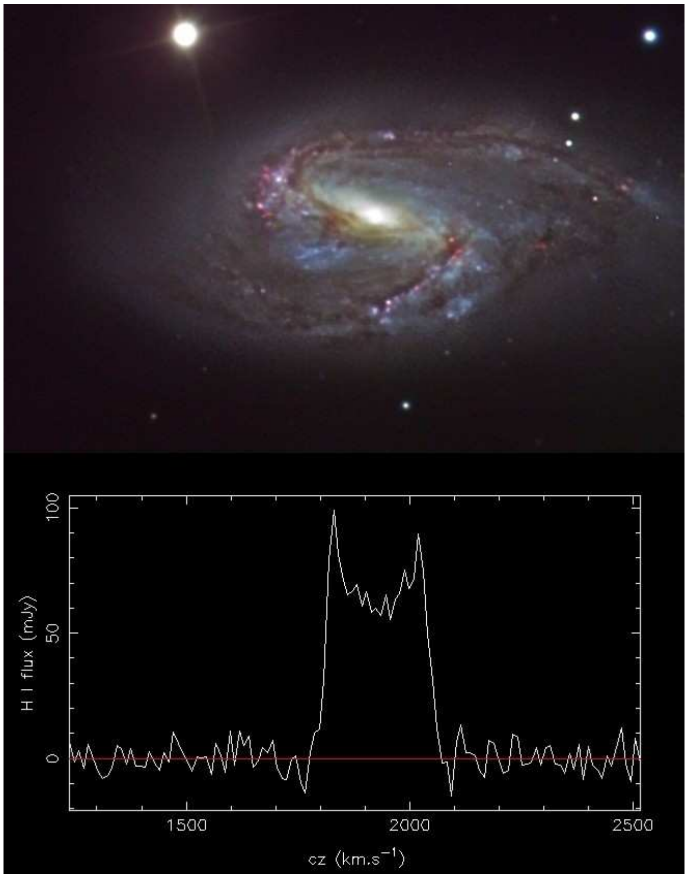
TUORLA OBSERVATORY
TURKU UNIVERSITY
VÄISÄLÄNTIE 20 - 21500 PIIKKIÖ
FINLAND

Training period report
Kinematics of the Local Universe and Cosmology: An Odyssey from Hydrogen spectra to the Hubble constant

Pierre GUILLARD

Supervised by Mikko Hanski, Gilles Theureau and Pekka Teerikorpi

11th, July 2004



SECOND YEAR OF THE “MAGISTÈRE” PROGRAM

ACADEMIC YEAR 2003-2004

Picture of the front page: photomontage of an optical image (above) of the M66 spiral galaxy by Robert Gendler and 21-cm line emission of the neutral hydrogen (H I) in the disc of the PGC 12876 galaxy.

Contents

Acknowledgements **7**

Abstract **9**

Introduction **11**

1 Fundamental concepts **13**

1.1 Link between cosmological distance scale and galaxies . . . 13
1.1.1 Hubble law and distance indicators . . . 13
1.1.2 First meeting with the Tully-Fisher relation . . . 13
1.1.3 Galaxies as a cosmological tool . . . 14
1.2 The KLUN galaxy survey . . . 15
1.2.1 The project . . . 15
1.2.2 The KLUN galaxy sample . . . 16
1.3 Tully-Fisher relation and Malmquist bias . . . 17
1.3.1 The Tully-Fisher relation in more detail. . . . 17
1.3.2 . . . and the Malmquist bias appears. . . . 17
1.3.3 . . . but the normalized distance method strikes back ! . . . 18
1.4 Problematics and purpose of this training period . . . 19
1.4.1 General context . . . 19
1.4.2 To understand and bypass the drawbacks of the TF relation . . . 20
1.4.3 The data . . . 20

2 Analysis of the H I spectra **21**

2.1 Characteristics of a H I spectrum . . . 21
2.1.1 Doppler effect, shape of the spectrum and inclination of a spiral galaxy . . . 21
2.1.2 Standard measurements on the spectrum : definitions and use . . . 22
2.2 Classification of the spectra . . . 23
2.2.1 “Zoology” of the H I profiles : examples in the jungle of the spectra . . . 23
2.2.2 Errors in the data and new analysis of the H I profile sample . . . 24
2.2.3 Quality criteria and classification system . . . 25

3 Tully-Fisher applications: on the way to H_0. . . **29**

3.1 An overview of the method . . . 29
3.1.1 Selection of the galaxies and completeness of the sample . . . 29
3.1.2 From the TF relation to H_0 in short . . . 30
3.2 First study on the totality of the TF diagrams . . . 30
3.2.1 TF diagrams for all the selected galaxies of the sample . . . 30
3.2.2 TF parameters: slope, zero point and scatter . . . 30
3.3 Investigation of the unbiased part . . . 32
3.3.1 The normalized distance method in practice . . . 32
3.3.2 Comparative study of TF parameters for the unbiased galaxies . . . 34
3.4 Towards an assessment of the Hubble constant . . . 35
3.4.1 Calibration and correction for the Malmquist bias . . . 35
3.4.2 Distributions of H_0 . . . 37
3.4.3 Comparative study of the values of H_0 and σ_{H_0} . . . 39
3.4.4 Uncertainties on the Hubble constant itself . . . 40

Conclusion **41**

A Glossary **43**

B Additional tables **45**
B.1 Detailed results of the spectral classification for the 3 sub-samples 45
B.2 Starting values of the TF parameters for the iterative normalized distance method 45

Bibliography **46**

List of Figures

1.1 The Hubble sequence of galaxies, [8] 14
1.2 Rotation curve for the galaxy NGC 2403, Begeman 1987 15
1.3 Map of the KLUN+ galaxies 16
1.4 Illustration of the second Malmquist bias and the normalized distance method 19
1.5 Map of the 3069 KLUN+ galaxies used in this study 20

2.1 Schematic principle of the Doppler broadening of the 21-cm H I spectral emission line. 21
2.2 Standard measurements on the H I spectrum of the PGC 12876 galaxy. 22
2.3 Sample of 4 H I spectra illustrating the variety of the shapes. 24
2.4 Sample of 4 H I spectra illustrating some errors in the earlier data analysis 25
2.5 Sample of 6 H I profiles, showing the different spectral classes. 26

3.1 Tully-Fisher diagrams for all the selected galaxies, and all the classes. 31
3.2 Normalized distance diagrams and TF unbiased relations for I band. 33
3.3 Influence of the quality of the spectra on the unbiased TF slope, zero-point and scatter. 35
3.4 Illustration of the calibration: Tully-Fisher diagrams, unbiased and calibrated TF relations. 37
3.5 Histograms of the distributions of H_0 for the five bandpasses. 38
3.6 Influence of the quality of the spectra on the Hubble constant and the scatter of its distribution. 39

List of Tables

2.1 Sharing out of the spectra within the different classes. 27

3.1 The five wavelength bandpasses B, I, J, H and K 29
3.2 Tully-Fisher parameters for the whole diagram: slope, zero-point, scatter and number of points. 32
3.3 Unbiased Tully-Fisher parameters: slope, zero-point, scatter and number of points. 34
3.4 Hubble constant mean values and sigma of its distribution in each bandpass. 39
3.5 Hubble constant mean values and their uncertainties in each bandpass. 40

B.1 Sharing out of the spectra within the different classes for the three sub-samples. 45
B.2 Input Tully-Fisher parameters for the first loop of determination of the plateau. 45

ACKNOWLEDGMENTS

First and foremost, I would like to especially thank Mikko Hanski, with whom I mainly worked during these three months. It is difficult to ask more from a supervisor. I greatly appreciated his kindness, his dynamism, his availability, the clarity of his explanations, and his patience, despite my many questions ! We have always worked as a team, in a relaxed atmosphere. A wink to his wife Mari, for her kindness. Both welcomed me very warmly in Finland, and did everything there to make my life easier. Thanks also for your advices concerning sauna, food, museums, and finnish rock music !

I would like to acknowledge Gilles Theureau and Pekka Teerikorpi. I particularly thank Gilles for having get me in contact with his finnish colleagues. From France, either by phone or, more often, by e-mail, he never denied precious suggestions, interpretations and recommendations. During his visit in Finland, we made big steps forward concerning some remaining obscure points. I thank Pekka for his advices, clarifications, and huge skills in cosmology. I am also very grateful to him for the financial help he managed to get for me. I am very honoured to have had such a team of scientists on my back !

I also appreciated the hospitality of the Tuorla observatory staff. Relaxing coffea breaks (thanks Mari and Mikko for the translations of some conversations ...!) and good sähly[1] or soccer games also contributed to this friendly work atmosphere. A special acknowledgement to Kari Nilsson for having allowed me to spend two complete nights with him observing quasars, just for the pleasure to use a one-meter optical telescope ! I would also like to thank Pekka Heinämäki for having allowed me to follow the cosmology courses of the *Planck summer school*. Now I understand better how one can do pre-big bang cosmology with the string theory...!

Thanks Luca, Paula, Eva, Janne, Stefano, Liisa and Luisa for your friendship, for being easily persuaded into going out for a walk, a concert or a barbecue, and for the good time spent in Turku.

A thought to my former supervisor, Karl-Ludwig Klein, for always having kept in touch with me, and without whom this training wouldn't have been possible.

I will always be grateful to my family and friends, especially those who sent me a lot of mail during these three months, and last but not least, a big kiss to Carole, for her love, support and understanding.

---

[1] Finnish national sport, sähly is an indoor floorball, each player having a stick.

ABSTRACT

The main topic of my master training period belongs to the KLUN program for measuring the kinematics of the local universe and observing its cosmological structures, in which the problem of extragalactic distance scale plays a fundamental part.

The direct Tully-Fisher relation, a correlation between absolute magnitudes and rotational velocities of spiral galaxies, is used as an distance indicator. In order to determine the parameters of this relation, one must calculate the rotation speeds of a large number of galaxies, which are derived from the study of neutral hydrogen (H I) spectra mostly made at the Nançay Radiotelescope.

As I noted errors in some of the published values of the central velocity and linewidths of the 21-cm H I spectral lines of the Nançay archive, I recalculated these parameters from a sample of 3312 spectra. Then, in order to reduce the important scatter shown by the Tully-Fisher relation, these spectra have been classified according to their quality and shape, and a sub-sample of very reliable data has been extracted.

A unique comparative study of the scatter and the Tully-Fisher parameters has been done, with respect to the classification of the data. Then, I applied the Tully-Fisher relation for deriving the distances of the selected galaxies. To this purpose, I paid a special attention to the understanding and treatment of the biases in the distances that appears when one works with an observed sample of galaxies. Finally an assessment of the Hubble constant is given, as well as its uncertainty.

Introduction

During my training period, I was working at the Tuorla Observatory, in Finland. Founded in 1952, the observatory is part of the Väisälä Institute for SpacePhysics and Astronomy (VISPA), formed at the University of Turku in 2001. It is now a growing independent research institute, in which about fourty people are working, mostly on the following fields:

- Cosmology : large scale structure of the universe, blazars, cosmic evolution of quasar host galaxies, extragalactic distance scale…
- Optics : MAGIC telescope collaboration, construction of the 3.5 meter Herschel-satellite mirror (ESO)…
- Dark matter : massive black holes, giant molecular clouds, white dwarfs…
- Stars : X-ray binaries, K-dwarfs, production of helium in the cosmos, solar activity, energetic solar particles…

The general framework of this training is the determination of distances in astronomy, which is becoming a red line in the history of observational cosmology. The estimation of distances is more tricky than it seems, but this is one of the most fundamental problems in astronomy. To simply know the scale of the cosmos, to assess the Hubble constant H_0 that measures the expansion rate of the universe, or to study its kinematics, we have to determine the distances and the velocities of the galaxies. To this purpose, we use the detection of the neutral hydrogen emission line in galaxies, at radio wavelengths, and the so-called Tully-Fisher relation.

The outline of this report is such that the physical notions and the tools for measuring the extragalatic distance scale are discussed in chapter 1. Chapter 2 describes the analysis of the observational data required for using these tools. Applications and results are given in chapter 3.

Note that the definition of the underlined words can be found in the glossary, appendix A.

Chapter 1

Fundamental concepts

This chapter presents not only the theoretical basis required for the understanding of the following parts, but also a description of the research project in which I was involved, and finally a brief inventory of some open questions and difficulties we have to face.

1.1 Link between cosmological distance scale and galaxies

1.1.1 Hubble law and distance indicators

Basically, cosmology aims at understanding and describing the origin, the structure and the fate of the universe. Of course, the precise estimation of cosmological distances plays an important part in this purpose. The Hubble law itself formulates the problem in simple terms. Indeed, this linear relation states that the expansion of the universe causes the recession velocity of a galaxy $V = c\,z$, where z is the redshift and c the speed of light, to be proportional to its distance d from the observer[1] :

$$V = H_0\,d \tag{1.1}$$

By the way, if one knows the value of the *Hubble constant* H_0, one can derive d from the single measurement of the velocity of the galaxy, and that is why this law is such a powerful tool for estimating extragalactic distances[2]. But the uncertainty on H_0 is still huge, since its value seems to be between 50 and 80 km.s^{-1}.Mpc^{-1}. To estimate it, one needs to measure the distances of galaxies independently of the redshift, which is much more difficult than deriving the recession velocity from the shift of a spectral emission line for instance.

To deduce the distance of a relatively nearby galaxy, we use some distance indicators, so-called *"standard candles"*. These standard candles can be Cepheid variable stars for instance, which really mark out the distance ladder. The first step is to derive the distances to Cepheids in our galaxy from parallax measurements. According to the definition[3] and the principle of uniformity of physical laws, a distance indicator is a group of objects having the same properties from one galaxy to another. Hence, this means that Cepheids in another galaxy have the same properties than those in the Milky Way. So we can calibrate their period-luminosity relation in our galaxy, and use it to deduce the distances to Cepheids in the other galaxy.

However, one can only apply the calibration of the standard candles, also called *the primary distance indicators*, to a local volume of space. Indeed, beyond about 20 Mpc Cepheids become too faint, and so alternative means, i.e. *secondary distance indicators*, are needed if we want to estimate the distances of a large, remote sample of galaxies.

1.1.2 First meeting with the Tully-Fisher relation

One of the most popular secondary method is based on the correlation between the luminosity of *spiral galaxies* and their rotational velocities. This is known as the Tully-Fisher relation (hereafter **"TF relation"** for short), described with more details in section 1.3, and firstly used for measuring extragalactic distances by Tully and Fisher in 1977.

The general use of the TF relation can be summarized in a simple way as follows : given the calibration of the TF relation, the rotational velocity allows to predict the absolute magnitude. The distance is then deduced from the absolute and apparent magnitudes.

---

[1] Note that here z is not due to a Doppler effect, but to the growth of the space itself. Moreover, the non-relativistic formula $V = c\,z$ is valid at small redshifts, i.e. $z \ll 1$.

[2] Usually measured in Mpc, where 1 Mpc = 1 Megaparsec $\simeq$ 3.10^{22} m

[3] Lundmark, 1946.

One must be aware of some selection effects, however : from large distances, we only see the most luminous objects, and so the average luminosity of our sample changes with the distance. The result is that the apparent TF relation differs from the true, physical one. This issue is discussed in section 1.3.2.

1.1.3 Galaxies as a cosmological tool

- **Basic morphological classification of galaxies**

  In a simplistic manner, galaxies can be splitted into three main morphological types, according to Hubble's classification (1936) shown on figure 1.1:

  - **Spiral** galaxies have three main components. Firstly, the *bulge*, a spherical structure found in the center of the galaxy, mostly contains older stars. Secondly, the *disk*, made up of dust, gas and younger stars, forms arm structures. Thirdly, the *halo*, contains old clusters of stars[4], but also the dark matter. Note that spirals galaxies are classified into two groups, ordinary (denoted S) and barred (SB), and further divided to different morphological types (*Sa*, *Sab*, *Sb*,...) as the figure 1.1 suggests.
  - **Elliptical** galaxies are shaped like an elongated sphere, and classified according to their ellipticity. They have no particular axis of rotation, and do not usually rotate, that is why they are not useful for the TF relation.
  - **Irregular** galaxies, not shown is the figure 1.1, have no particular or symmetrical structure.

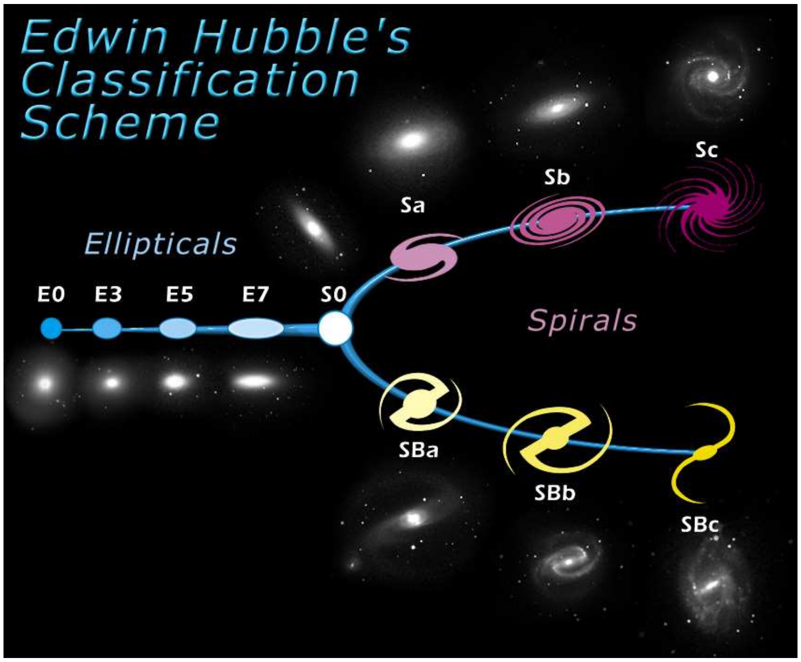


Figure 1.1: The Hubble sequence of galaxies, [8]

Since the rotation velocity is one of the physical quantities we want to measure for the TF relation, the two following paragraphs aim at precising what this velocity actually is and how we can measure it.

- **The rotation curve of a spiral galaxy**

The rotation curve of a galaxy is the evolution of the orbital circular velocity $V_{rot}(R)$ of stars or gas clouds with respect to R, the distance from the center. This variation reflects the distribution of mass in the galaxy. Most of the galaxies have rotation curves that show solid body rotation in the very center, which means that $V_{rot}(R) \propto R$, following by an almost constant rotation velocity in the outer parts. This typical shape can be seen in figure 1.2.

However, if one considers the mass $\mathcal{M}(R)$ within the radius R in circular orbit in a galaxy and writes the equilibrium between gravity and centrifugal acceleration,

$$\frac{G\,\mathcal{M}(R)}{R^2} = \frac{V_{rot}^2(R)}{R}, \tag{1.2}$$

---

[4] known as *globular clusters*

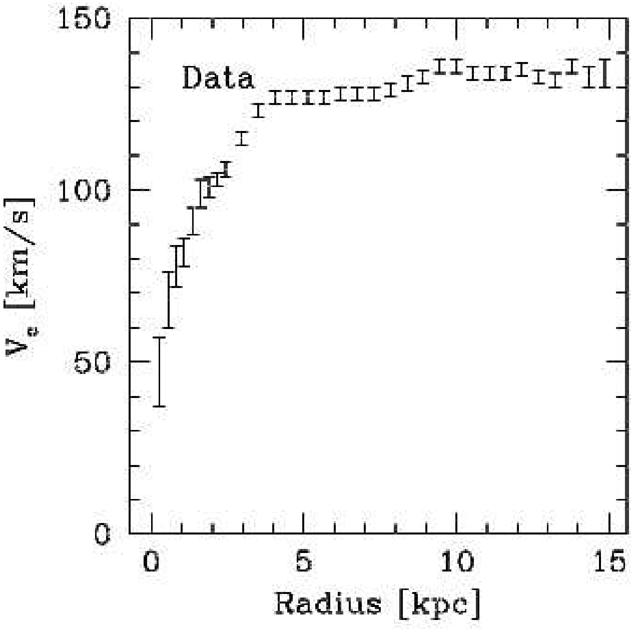


Figure 1.2: Rotation curve for the galaxy NGC 2403, Begeman 1987

one would expect, at large R, when most of the mass in concentrated at the center, a "Keplerian decline" i.e. $V_{rot}(R) \propto R^{-1/2}$. Thus we need to add further mass in the system, and in reality the flatness of the second part of the rotation curve is a strong evidence for the presence of dark matter in the halo.

The quantity used in the TF relation is the maximum of this rotation velocity, hereafter V_{max}. Usually, this rotational velocity is measured via the Doppler broadening of the neutral atomic hydrogen emission line, at the wavelength of 21 cm, since spiral galaxies are easily detected in this spectral line using radiotelescopes.

- **The 21-cm emission line of the neutral atomic hydrogen (H I)**

Hydrogen represents 90% of the chemical composition of the universe, and in the interstellar medium we find it mostly in its neutral atomic (H I) or molecular forms. In 1944, Van De Hulst predicted this so-called *21-cm H I line*, emitted by Hydrogen atoms when the relative orientation of the spin of their electron and proton with respect to each other changes. This happens when eventually Hydrogen atoms collide and the spins are aligned in a parallel state. But then, eventually, the electron can flip its spin to get in lower energy state, i.e. the anti-parallel one. Consequently, a low energy photon, at the radio-wavelength of 21 cm, is emitted[5].

This is a rare transition[6], but because of the huge amount of H I, this line is very strong, which is the first advantage to use the radio measurement. Moreover, the H I envelope of spirals galaxies extends much further out than the optical emission, thus it provides a unique mean of determining the mass distribution in their outer regions. In other words, this extended distribution of H I regions is such that the flat portion of the rotation curve is well sampled, providing an accurate measurement of the maximum rotational velocity. As a general rule, the H I radio emission is the best tracer for the overall gas distribution and kinematics in spiral galaxies.

1.2 The KLUN galaxy survey

1.2.1 The project

Starting from 1988, the KLUN project, for Kinematics of the Local Universe, is a French-Finnish collaboration that gathers people from different institutes : The Paris-Meudon Observatory group (G. Theureau, L. Bottinelli, L. Gouguenheim, N. Durand, N. Hallet), the Lyon Observatory group (G. Paturel, C. Petit), and the Turku University group (P. Teerikorpi, M. Hanski, T. Ekholm).

In 1998, the KLUN team has built a catalog of 6 700 spiral galaxies (Theureau et al. 1998, [6]) having measurements of the physical parameters useful for the TF relation, as for instance the 21-cm H I line width. Since then, the KLUN+

---

[5] which corresponds to a frequency of 1420 MHz.

[6] on average every 10 million years for one atom...

project has been launched and aims at collecting further H I line width data. Last year, the number of galaxies in the KLUN+ sample was around 16 000, and these measurements are combined in the HYPERLEDA database[7] (G. Paturel et al. 2003, [9]) which includes the magnitude data of some whole sky surveys at different wavelengths.

1.2.2 The KLUN galaxy sample

A large part of the H I spectra have been measured by the Nançay Radiotelescope (France). The KLUN+ sample is the largest compilation of H I line measurements for galaxies, and the data are stored and homogenized to a standard system in the frame of HYPERLEDA (G. Paturel et al. 2003, [9]).

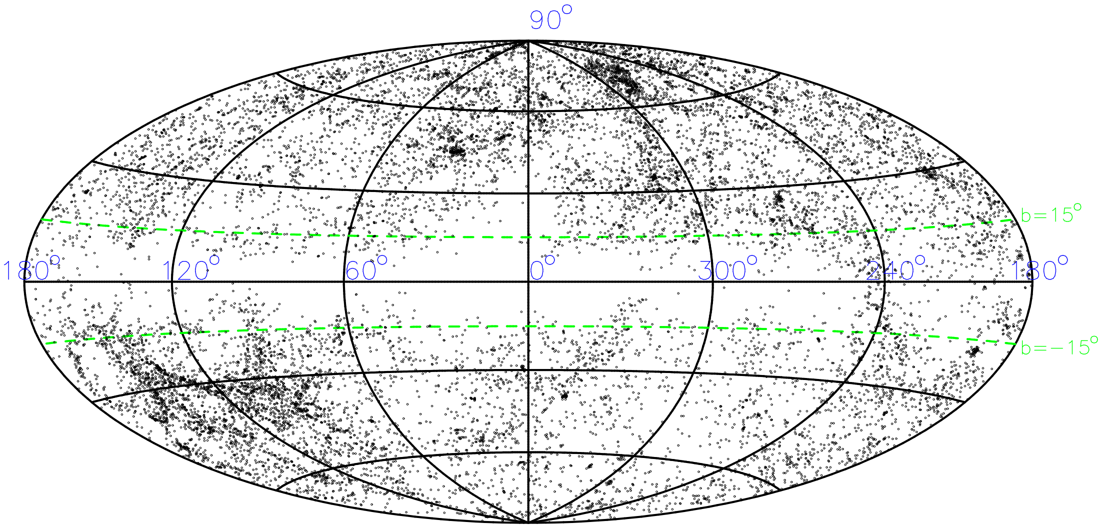


Figure 1.3: 15666 KLUN+ galaxies plotted on an aitoff projection of the sky in galactic coordinates. In the region between the green dashed lines, uncertainties on the magnitudes are important, because of the dust present in the plane of the Milky Way.

The figure 1.3 shows the map of the KLUN+ galaxies and the incomplete regions. Low galactic latitudes $|b| < 15°$ will be excluded for TF studies because of the obscuring dust lying in the disk of our galaxy. For a detailed description of the KLUN sample, see Theureau et al. 1997, [5].

Why do we need such a large sample of galaxies ? There are several reasons for that, and some of them will be enlightened later, but the main points are the following :

- Firstly, in order to have an homogeneous and complete catalog, we need to collect as many galaxies as possible. This completeness and dense coverage of the sky are required if one wants to study the structure and the kinematics of the local universe.
- Then, only a large number of objects allows to study the statistical properties of the sample itself. For instance, is there any bias due to some selection effect when one constructs the sample ? Moreover, the knowledge of these caracteristics is essential, if we want to correct a possible bias. This point will be discussed in section 1.3.2.
- We will see that the study of the TF relation itself requires a large and well understood sample, as well as its application on the determination of the mass distribution or the velocity fields in the galaxy clusters and superclusters.
- Finally, increasing the population of the sample allows to determine the Hubble constant H_0 on a large distance scale.

---

[7] Lyon Meudon Extragalactic Database, `http://www-obs.univ-lyon1.fr/hypercat/`

1.3 Tully-Fisher relation and Malmquist bias

1.3.1 The Tully-Fisher relation in more detail…

We now know that the TF relation links the absolute magnitude M to the maximum rotation velocity of the galaxy V_{max}, and the correlation is the following :

$$M = a_1 \log(V_{max}) + a_0 \; , \tag{1.3}$$

where a_1 is the TF slope and a_0 the zero point. The physical origin of this relation arises from the correspondence between mass and luminosity. A spiral galaxy rotates under the influence of its own gravity and so the rotational speed is connected with its mass. A more massive galaxy also contains more stars, and is therefore more luminous. If we assume a constant mass to luminosity ratio (Hanski 2001, [2]), one can rewrite equation 1.2 with the same notations :

$$V^2_{rot}(R) = \frac{G\,\mathcal{M}(R)}{R} \propto \sqrt{L(R)}, \tag{1.4}$$

where $L(R)$ is the integrated luminosity at the distance R from the center. Since the definition of the absolute magnitude is $M = -2.5\log(\frac{\mathcal{L}}{\mathcal{L}_\odot})$, where $\mathcal{L}$ and $\mathcal{L}_\odot$ are respectively the total luminosity of the galaxy and a reference luminosity[8], one would expect to obtain a TF relation with a slope $a_1 = -10$.

There are a few complications to this simple picture of the TF relation. Firstly, the total luminosity of a particular spiral galaxy results from its star formation history, and this might be expected to vary from galaxy to galaxy. Secondly, we now know that spiral galaxies have huge amount of dark matter, so the mass of a galaxy is not necessarily well represented by the amount of mass contained within its population of stars. In other words, the ratio of mass to light for spiral galaxies might vary at a given total mass.

However, the observed scatter of the TF relation is quite small, which means that galaxies are relatively homologous. The fact that the TF relation is well verified thus provides strong constraints on models for the formation and evolution of the spiral galaxies. So we can still measure galaxies distances through the study and the calibration of the TF relation, by using primary distance indicators (cf section 1.1.1).

But why the TF relation as a secondary distance indicator is so popular ? The main reason is that the universe offers us a large number of spiral galaxies, and moreover regularly distributed on a wide distance range[9]. Thus we have at our disposal an ideal sample of suitable objects for TF studies, like the assessment of H_0, or the calculation of the distribution of matter in the local universe.

1.3.2 …and the Malmquist bias appears…

The basic ideas of the systematic errors which appear when one tries to calibrate *standard candles* on a distance ladder and to estimate ”true” extragalactic distances, are presented in a simple manner. This point is a rich and tricky subject, but fundamental in astronomy, and in general called *Malmquist bias* whereas two main classes of biases exist. For a precise description of these statistical ”challenges”, see the review of P. Teerikorpi (1997, [1]).

- **”Bias of the first kind” or ”classical Malmquist bias”**

The first kind of bias is geometrical, due to a ”volume effect”. Simply, if we assume that the universe is homogeneous[10], which means that the space density distribution $\rho(r) = \rho$ is constant, the number of galaxies located at a distance $r \pm \frac{dr}{2}$ is $N(r) = 4\pi\rho r^2 dr$. So, there are more objects between r and $r + \frac{dr}{2}$ than between $r - \frac{dr}{2}$ and r.

Now let us consider a subsample of galaxies having the same *derived* distance d_{der}, or the same derived distance modulus, calculated from the apparent magnitude m and the absolute magnitude M_{TF} given by the TF relation: $\mu_{der} = m - M_{TF}$. These objects actually have a scattered distribution of *true* absolute magnitudes around M_{TF}. Because of the volume effect described above, there are consequently more objects brighter than M_{TF}, compared to fainter ones, so that the real average magnitude of our subsample is indeed brighter the assigned M_{TF}.

Hence, in our set of galaxies having $\mu_{der} = const$, there are more objects with true distances which are larger than the derived one. Thus, the maximum of the distribution of true distances is constantly higher than d_{der}. In other words, we can sum up this effect by saying that the classical Malmquist bias is the systematic average error in the true distance modulus for a class of galaxies with $\mu_{der} = const$ (Teerikorpi 1997, [1]).

How do we correct the bias of the first kind ? One assumes that the true distance modulus μ_{true} follows a Gaussian function[11] whose scatter is characterized by the parameter σ. As stated above, the “shift” of the mean of the distribution of true distances with respect to the derived distance only depends on the width of the distribution. So one can obtain the following correction, valid for a homogeneous space distribution:

---

[8] This reference luminosity $\mathcal{L}_\odot$ can be the luminosity of the sun for example.

[9] contrary to the elliptical galaxies for instance, which are mainly found in clusters.

[10] Homogeneity is a good approximation at large scales, i.e at distances larger than 100 Mpc.

[11] This assumption was made by Malmquist in 1920.

$$\langle \mu_{true} \rangle = \mu_{der} + 1.382\sigma^2. \tag{1.5}$$

Note that this formula reflects well the fact that, on average, the true distance of an object belonging to a $\mu_{der} = const$ class is larger than the derived one.

- **”Bias of the second kind” or ”selection bias”**

The second Malmquist bias is due to the statistical characteristics of the sample itself, when a set of objects is chosen in a flux-limited way. As an example, let us consider our KLUN sample : because we only see up to an apparent magnitude limit m_{lim}, we have here a set of galaxies which is complete to this given apparent magnitude. So, at a fixed *true* distance d^0_{true}, the distribution of absolute magnitudes is biased, that is to say it is cut at an upper magnitude limit given by $M_{lim} = m_{lim} - 5\log(d^0_{true}) - 25$, according to equation A.1.

Thus, when the distance increases, intrinsically faint galaxies are more and more underrepresented, relative to an unbiased (true) distribution of its absolute magnitudes. Moreover, at a fixed true distance, the observed mean luminosity of a fainter class of galaxies deviates more from the true average than for the brighter one. This effect can be clearly seen on the figure 1.4(a), section 1.3.3. On this plot, galaxies in class 1 ($\times$) are on average brighter than the ones in class 2 ($\circ$). One may also say that for class 2, the bias begins at smaller distances than for class 1. And the result is a systematic error in the distances derived from the TF relation. In terms of distance modulus, one can briefly define the second Malmquist bias as follows : it is the systematic error in the average of the derived distance modulus $\langle \mu_{der} \rangle$ for a class of galaxies with $\mu_{true} = const$ (Teerikorpi 1997, [1]).

The KLUN team has developed a special technique called the *normalized distance method* in order to overcome this bias. The basic principles are presented in the following section, and the application of this method will be illustrated later, in chapter 3.

1.3.3 . . . but the normalized distance method strikes back !

The principle of the method of normalized distances is based on an identification of an unbiased part of the M vs. $\log(d)$ diagram, common for all the rotation velocities classes. This is possible if one revalues the distances d to normalized ones d_n, so that the shift between the true and observed average of the absolute magnitudes (i.e the amount of bias) be the same at any fixed normalized distance. One can see the result of such a transformation of the distance axis by comparing figures 1.4(a) and 1.4(b).

After this new distance scaling, the average lines $\langle M_1 \rangle$ and $\langle M_2 \rangle$ in figure 1.4(b) start to rise at the same normalized distance $d_{n,lim}$. The so-called *unbiased plateau* can be identified, where $\langle M_{1,2} \rangle$ stay relatively close to the true average. It corresponds to an unbiased subsample from which we can evaluate the unbiased Tully-Fisher parameters (slope and zero point of the TF relation, equation 1.3).

How do we derive the normalized distance ? A brief analytical description of the basic method is given here[12]. One calls d_{lim_1} and d_{lim_2} the distances from which the observed average lines $\langle M_1 \rangle$ and $\langle M_2 \rangle$ start to get biased on figure 1.4(a). According to equation A.1, these limiting distances are given via : $5\log(d_{lim_{1,2}}) = m_{lim} - \langle M_{1,2} \rangle - 25$. So, for the two classes of galaxies having different rotational parameters $p_{1,2} = (\log V_{max})_{1,2}$, the two limiting distances are linked by the relation $d_{lim_2} = d_{lim_1} 10^{0.2(\langle M_1 \rangle - \langle M_2 \rangle)}$. In other words, these limiting distances differ by the factor $10^{0.2a_1(p_1-p_2)}$, according to the TF relation (equation 1.3). Hence, if one multiplies the distance d by the inverse factor, $\langle M_1 \rangle$ and $\langle M_2 \rangle$ will have the same behaviour along the thus defined normalized distance axis (Theureau et al. (1997), [5]).

Generally, one re-normalizes the distances of all the classes $p = \log V_{max}$ to a given p_0. So one can write now the analytical expression of the normalized distance as follows :

$$d_n = d\, 10^{-0.2a_1(p-p_0)}, \tag{1.6}$$

which is close to the original formula given in Bottinelli et al. (1986), [11].

But how do we make the correction or, in other words, how do we free the distance (derived from the TF relation) from the bias ? Clearly, the interest of having done such a new scale in the distances is that we can now calculate the amount of bias for all the galaxies at each normalized distance. This is done by assuming a gaussian distribution (σ) of the points around the observed absolute magnitude M, and the result is a function depending on three parameters : d_n, the scatter σ and m_{lim}. So we can add this quantity to the distance modulus in order to have the corrected one :

$$\mu_c = \mu + f(d_n, \sigma, m_{lim}) \tag{1.7}$$

As years went by, the KLUN team has made lots of refinements of this method. Some of these corrections will be point out in section 1.4.2, because they are linked with some progresses in the understanding of the TF relation itself.

---

[12] For a more detailed review, see Bottinelli et al. (1986), [11] and Theureau et al. (1997), [5].

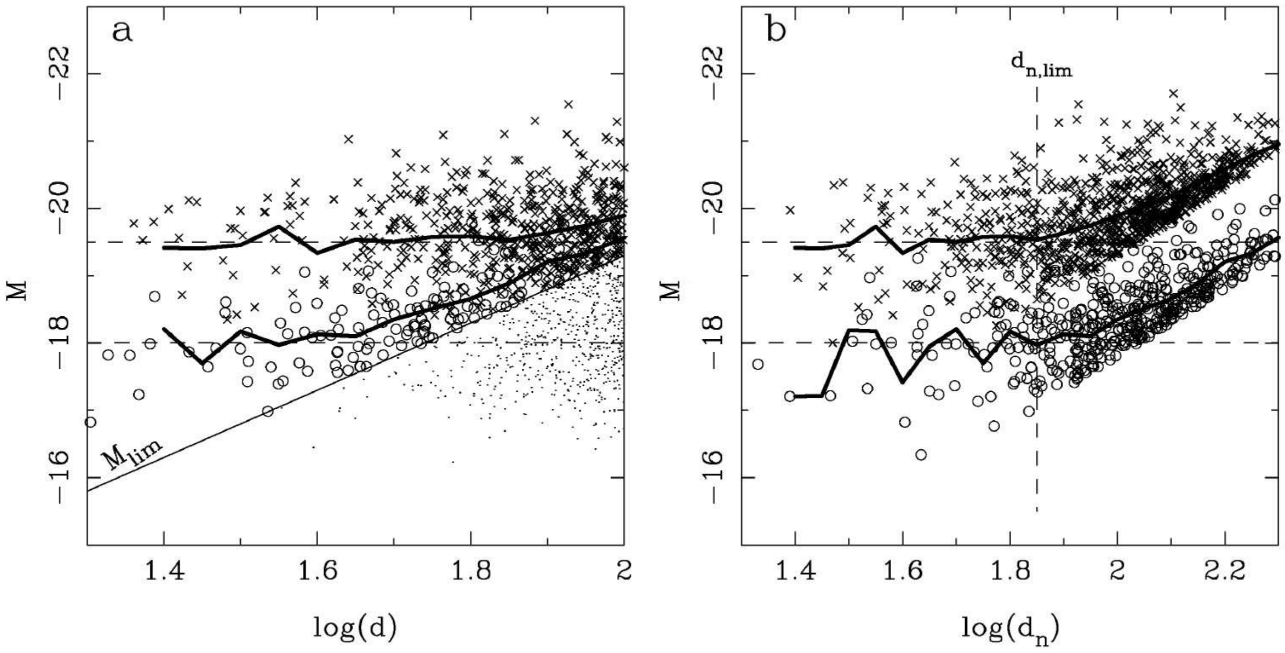


Figure 1.4: On both plots are represented two sets of galaxies (1: × and 2: ○) with different fixed rotational velocities, with $(\log V_{max})_1 > (\log V_{max})_2$, so $\langle M_1 \rangle < \langle M_2 \rangle$. Dashed and thick continuous lines show respectively the true and observed averages of $\langle M_{1,2} \rangle$ at different distances.
Fig. (a) illustrates the second Malmquist bias. Galaxies below the magnitude limit line (M_{lim}) are not observed, so they are marked with dots. The bias (i.e. the rise of the $\langle M_{1,2} \rangle$ lines) is a direct consequence of the cut in magnitudes.
Fig. (b) suggests the principle of the normalized distance method. Distances are normalized (d_n) so that the bias begins at the same distance $d_{n,lim}$ for the different classes. *Unbiased plateau* is the $d_n < d_{n,lim}$ region. From Hanski (2001), [2]

1.4 Problematics and purpose of this training period

After a general presentation of the framework, more precise objectives of this training and a brief description of the data I made use are given in the three following sections.

1.4.1 General context

The first step was to understand the physical and statistical basis presented above, to be aware of what we know, what we don't know, and what are the aims of the KLUN studies. Very briefly, the main goals are:

- To understand the structure of the universe by measuring and mapping the kinematics of the local galaxies. To study the repartition of matter in the universe, and the radial distribution of galaxies. To probe the internal dynamics and evolution of the galaxies.

- To use the properties of the local universe (where one can make precise measurements) for the comprehension of the origin and dynamics of large scale structure and the role of dark matter. To compare the observed dynamics and formation of the galaxy concentrations to theoritical models. To realise numerical simulations in order to model cosmological structure formation.

- To measure the expansion rate of the universe, via the assessment of the Hubble constant H_0.

If we want to succeed in these attempts, one must do a preparatory work, concerning the extragalactic distance scale, the comprehension of the biases mentioned above and the study of the Tully-Fisher relation, which is the general framework of my training.

1.4.2 To understand and bypass the drawbacks of the TF relation

As we have just seen before, the determination of extragalactic distances is a problem in itself ! First, one must collect a large and complete sample in order to use the TF relation, correct the derived distances from the Malmquist biases, and then calibrate the TF relation in order to finally obtain the absolute distances. But this is not all !...

The TF relation itself suffers from a important drawback : its scatter. The main reasons are the observational errors, the uncertainties when one takes into account the inclination[13] of the galaxy and the extinction of the signal by the dust lying in our galaxy, and the intrinsic scatter of the relation. All this causes errors in the derived TF distances of about 20%. Of course, one goal of the KLUN studies is to reduce these uncertainties on the distances. But how ?

Concerning the measurements of the apparent magnitudes, KLUN galaxies are chosen so that the uncertainties on m are not too high[14]. One can also choose a measurement in a wavelength band which is less affected by the extinction of the dust, like in near infrared. But another way to reduce errors in the distances, is to study the intrinsic scatter of the TF relation, and to refine it in order to reduce the uncertainties. For instance, it has been shown that there is a morphological type dependence in the TF relation (Theureau et al. (1997), [4]). By selecting one morphological type of galaxies, one can reduce the TF scatter. Besides, this type dependence has lead to a refinement of the normalized distance method, and in particular to a modification of equation 1.6 (Theureau et al. (1997), [5]).

1.4.3 The data

I had at my disposal during this training a set of 3312 H I spectra measurements made in Nançay, which corresponds to a sample of 3069 different galaxies effectively detected, meaning that one galaxy can have several H I measurements. A map of these galaxies is given figure 1.5. The empty region on this map corresponds to the limit in declination under which the Nançay radiotelescope cannot go, i.e. $\delta < -38°$.

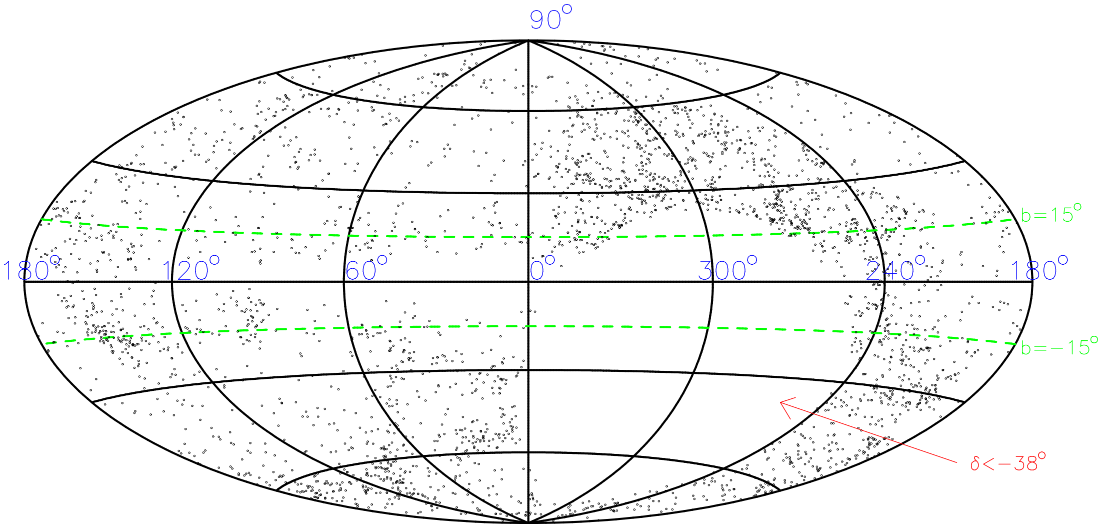


Figure 1.5: Map of the 3069 galaxies of the sample used in this study, plotted on an aitoff projection of the sky in galactic coordinates.

This sample is actually a collection of three sub-samples called REF023289, REF022359, and REF099999, corresponding to three previous publications: Theureau et al. 1998 [6], Paturel et al. 2003 [9] and Theureau et al. 2004 [in press].

The idea is to do a study of the TF relation and to try to reduce its scatter by a careful analysis of this data, and especially by a classification of these spectra. What does a H I spectrum look like and where does its shape come from ? How do we derive the rotation velocity from the H I line width ? How can we extract the best quality data in order to reduce the TF scatter ? All these points will be clarified in the next chapter…

---

[13] The inclination i is the angle between the normal to the plane of a galaxy and the line of sight

[14] Typically, the selection criterion is $\Delta m < 0.3$ mag.

Chapter 2

Analysis of the H I spectra

2.1 Characteristics of a H I spectrum

This part is dedicated to the measurement of the maximum rotation velocity of a spiral galaxy (V_{max}) from the spectrum of the 21-cm H I line, and to the physical basis behind this.

2.1.1 Doppler effect, shape of the spectrum and inclination of a spiral galaxy

As stated above, the rotation velocity is deduced from the width of the H I emission line. But why this line is widened ?

Basically, due to the spin of the galaxy, an observer will see part of the galaxy approaching him and part of the galaxy moving away from him. Figure 2.1 illustrates the fact that the rotation of the galaxy causes a special shape for the H I spectrum. Due to Doppler effect, neutral hydrogen emission in the spiral arms that are going away from the observer is redshifted, and blueshifted for those that are coming towards him. The motion will thus cause the H I line to be smeared out and to appear broad to the external observer.

The broader the line, the faster the galaxy must be spinning, and, according to the TF relation, it must be brighter. Because the rotation curve of the galaxy is flat, all the photons are blue or red-shifted with the same quantity. Then, the projection effect of the rotation velocity on the line of sight causes the typical double horn profile shown in figure 2.1. Note that the redshift z of the galaxy can be read in the center of the line. It corresponds to the global shift of the H I spectral line due to the radial velocity with respect to the observer (cf. section 1.1.1).

However, there are lots of complications to this simple view. Firstly, one must correct the line width by taking into account the inclination of the galaxy. Indeed, the Doppler broadening of the H I line only reflects the measurement of the projection of the hydrogen gas' motion on the line of sight. How do we determine the inclination angle i of a galaxy ? One simply describes a spiral galaxy like a flying saucer, or more exactly[1] like an ellipsoid, characterized by its ellipticity $\varepsilon = 1 - \frac{b}{a}$, where a and b are the major and minor axis diameters obtained from photographic data[2]. The inclination can thus be calculated from the projected shape

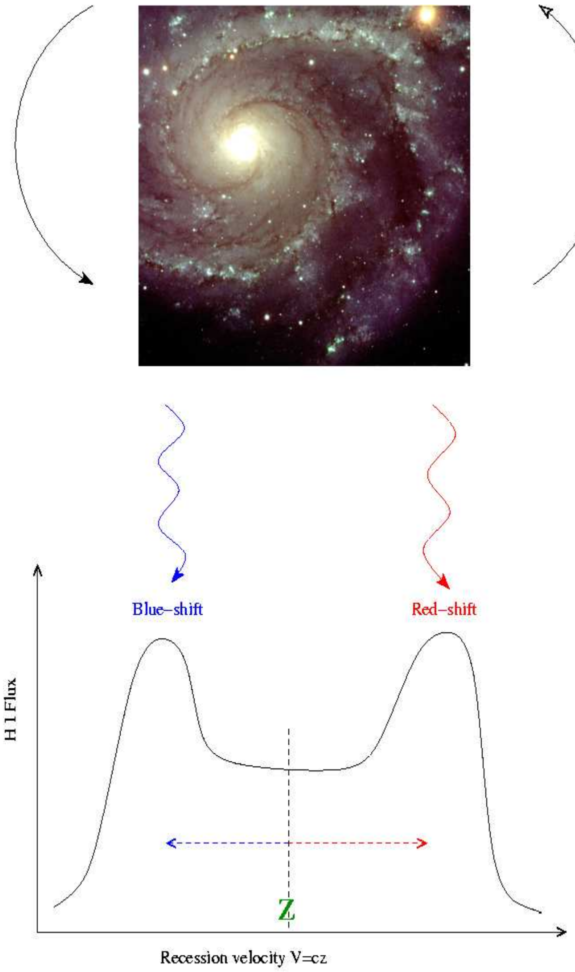


Figure 2.1: Schematic principle of the Doppler broadening of the 21-cm H I spectral emission line.

---

[1] although the following description is less poetic...

[2] In practice, thanks to modern CCD obervations, we measure the ellipticity from examining the contours of equal surface brightness (isophotes) in a galaxy image.

of the disk : specifically,

$$\cos^2 i = \frac{(b/a)^2 - \alpha^2}{1 - \alpha^2}, \tag{2.1}$$

where b/a is the observed axial ratio of the best-fitting elliptical isophote and $\alpha \approx 0.2$ the intrinsic axial ratio exhibited by an edge-on spiral galaxy ($i \simeq 90°$). One point must be emphasized : formula 2.1 is an approximation[3] valid for high inclinations, so in the KLUN sample, galaxies with $i < 32°$ are rejected. On the other hand, it is easy to see that for a face-on galaxy, the Doppler broadening of the H I line will be weak, so the rotation velocity is poorly determined. Indeed, if $i \simeq 0°$ the projection of the rotation velocity is very small, and that is why we also exclude face-on galaxies from our TF samples.

Then, we will soon see in section 2.2 that the quality and the shape of the H I line can vary a lot from one spectrum to another. Some explanations for such a variety in the spectra will be evocated. So how to define the H I line width in a reliable way ?

2.1.2 Standard measurements on the spectrum : definitions and use

The quantities measured on a H I spectrum have been standardized, so that one uses the so-called $\mathbf{W_{50}}$ and $\mathbf{W_{20}}$ i.e the widths at 50% and 20% level of the maximum of the line. An example of the the quantities we measure on the spectrum is given figure 2.2. The widths thus correspond to the “distance” separating the two external points of the profile at these intensity levels. The flux unit is the Jansky[4] (Jy).

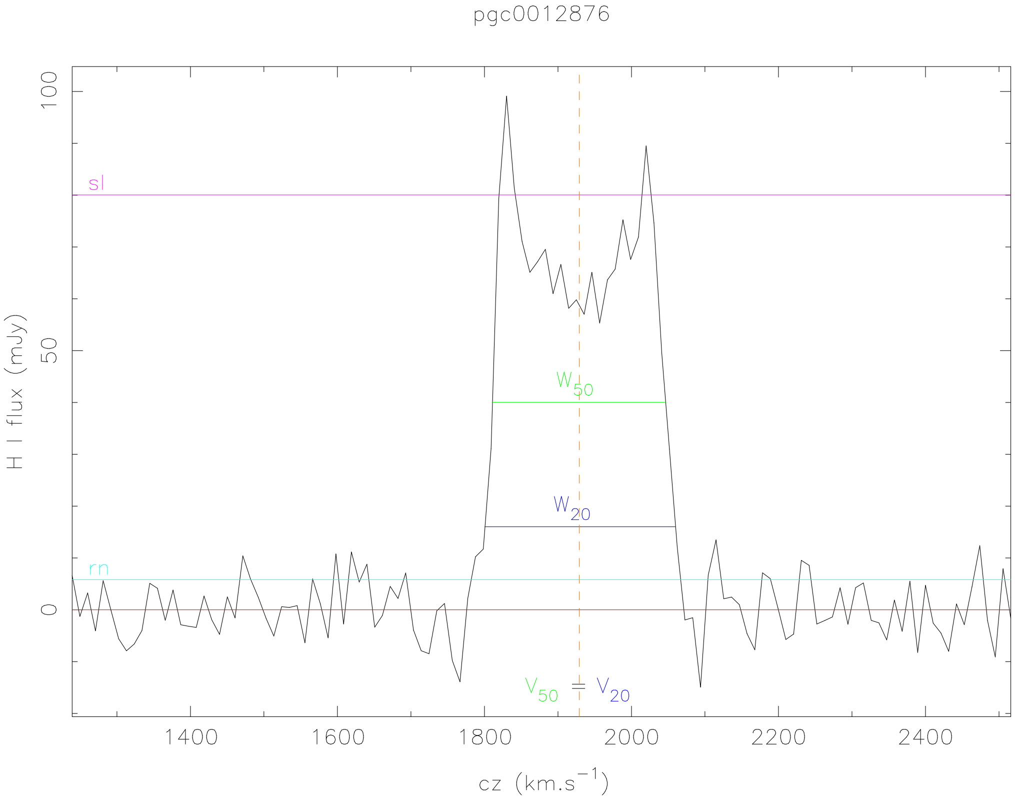


Figure 2.2: Standard measurements on the H I spectrum of the PGC 12876 galaxy.

First of all, the continuum component of the signal is substracted, so that the average of the noise be at the zero level (red line on figure 2.2). Then, if one wants to estimate the maximum of the line, so-called signal level **sl**, one must take into account the amount of raw noise **rn** (light blue line on figure 2.2). **rn** is simply the average of the positive part of the flat signal, after the spectrum has been “cleaned up” of the parasites[5]. To estimate accurately **sl**, we usually take the absolute maximum of the flux, and we substract the noise **rn**. Finally, knowing **sl**, we can measure directly the width of the line at the standard levels 20% and 50% of that maximum.

---

[3] It assumes that all spiral galaxies can be represented as a spheroid with the same intrinsic flattening. Complications arise from the fact that some galaxies do not have perfectly circular isophotes at face-on orientation, but this is probably the best we can do ! (Willick 1999, [12])

[4] Karl Jansky was the first to detect radio frequency emission from the Milky Way. 1 Jy = 10^{-26} W.m^{-2}.Hz^{-1}

[5] These parasites are essentially caused by terrestrial emissions and electronic noise.

To estimate the recession velocity $V = cz$ of the galaxy, one takes the abscissa of the middle of the line widths, so-called $\mathbf{V_{50}}$ and $\mathbf{V_{20}}$. On the spectrum figure 2.2, these two velocities are equal, and represented by the orange dashed line. However, in case of an asymetric H I line, $\mathbf{V_{50}}$ differs from $\mathbf{V_{20}}$, so one takes the average of the two. The radio measurement of the recession velocity of a galaxy is thus given by :

$$V = cz = \frac{\mathbf{V_{50}} + \mathbf{V_{20}}}{2} \tag{2.2}$$

For TF studies, we are also interested in the (maximum of the) rotation velocity V_{max}. We now know that the Doppler broadening of the signal is the direct consequence of the rotation of the galaxy, i.e. the width of the H I line represents roughly two times the rotation velocity. However, because of the galaxy inclination, the observed line width is smaller, by a factor $\sin(i)$, than the intrinsic value. Finally, if we define the width of the spectral line as the average between $\mathbf{W_{50}}$ and $\mathbf{W_{20}}$, the rotation velocity (corrected for inclination) and the radio line widths directly measured on the H I spectrum are linked by $2V_{max} = \frac{\mathbf{W_{50}}+\mathbf{W_{20}}}{2\sin(i)}$. According to the TF relation (equation 1.3), we rather calculate the log of V_{max}:

$$\log(V_{max}) = \log\left(\frac{\mathbf{W_{50}} + \mathbf{W_{20}}}{4}\right) - \log\left(\sin i\right). \tag{2.3}$$

One must be aware of a few details, however : Firstly, not only the levels of measurement (50% and 20%) but also the definition of the line width itself are quite arbitrary. Difficulties in the evaluation of the width arise when the spectrum has not a sharp and symetrical shape for instance. Then, when one measures the H I spectrum, one detects also the random motion of the neutral hydrogen gas in the galaxy. This creates an internal dispersion in the rotational velocities, which causes a further modification in the spectra. In other words, the radio line width does not only reflect the exact rotation velocity. Nevertheless, this random component is small compared to the global rotation, and is actually of the order of the resolution in velocity. Thus, we can still use this formula for TF studies.

2.2 Classification of the spectra

In this section we enter more into the details of the analysis of the spectra. Some errors in the previous measurements are pointed out, as well as the improvements made in the classification of the spectra.

2.2.1 "Zoology" of the H I profiles : examples in the jungle of the spectra

All the spectra of the sample I have made use (briefly presented in section 1.4.3) had been previously analysed, i.e the values of all the parameters mentioned in section 2.1.2 had been determined. Figure 2.3 illustrates how the shape of a profile can vary from spectrum to spectrum. Let us give a brief description of these four spectra (figure 2.2) and evoke some of the main reasons that might explain their look. One numbers the spectra from left to right, then from up to down.

1. The first spectrum of PGC 29 993 is one of the most beautiful[6] in this sample, which means that its shape is the typical double horn profile presented in section 2.1.1. The signal is intense and symmetrical. Hence, the line widths $\mathbf{W_{50}}$ and $\mathbf{W_{20}}$ are very accurate, as well as the measurement of the recession velocities $\mathbf{V_{50}}$ and $\mathbf{V_{20}}$ (vertical green and blue lines), since we have $\mathbf{V_{50}} \simeq \mathbf{V_{20}}$. This galaxy has an inclination angle $i = 64°$, which fulfils the criterion mentioned in section 2.1.1. Of course, this kind of H I profile is the best data for TF application.

2. The second one (PGC 64 939) presents a quite strong but asymmetrical profile. It seems that one detects essentially one spiral arm. There are various explanations, the galaxy may not have a symetrical structure[7], so the distribution of neutral hydrogen gas is not uniform for instance. Because of the difference between $\mathbf{W_{50}}$ and $\mathbf{W_{20}}$, the assessment of the rotation speed is not very accurate.

3. The third signal is a very intense single peak. The Doppler broadening of the line is weak, which means that the observed rotation velocity is small. As in the previous case, we cannot fully trust the V_{max} derived from this smooth profile. This kind of spectrum can be due to a face-on galaxy or a dwarf galaxy[8], but also to a partial detection of one galaxy that may overlap the field of view of the radiotelescope. These possible *confusions* in the detection are due to the poor angular resolution at the wavelength of 21 cm.

4. The last spectrum (PGC 558) is quite noisy, but still allows to measure the redshift and the line width. This kind of profile is quite common, due to the faintness of some objects. Of course there are worse spectra...

---

[6] Beauty being a subjective notion, this value judgment is entirely personal...

[7] For instance, gravitational interactions between 2 galaxies (or more) can lead to some deformations of the galaxy structure.

[8] Dwarf galaxies are typically small, low in mass and density. Most of them have also an irregular morphology, and have almost no rotation.

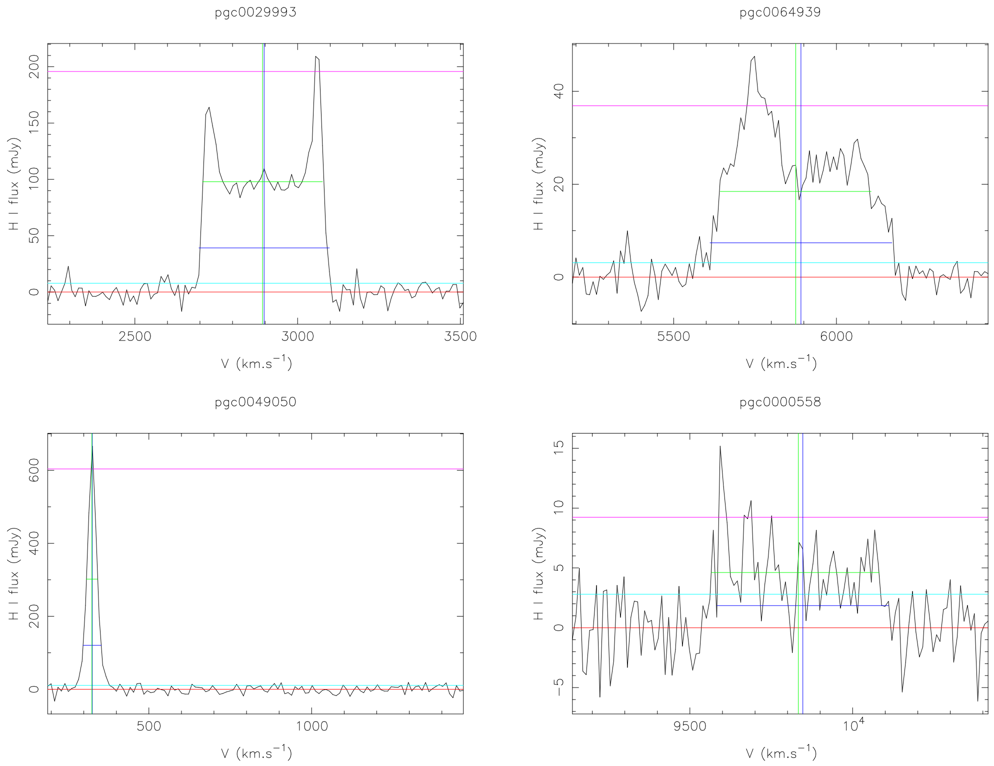


Figure 2.3: Sample of 4 H I spectra illustrating the variety of the shapes. The standard measurements done before this study are indicated, with the same color caption than figure 2.2.

2.2.2 Errors in the data and new analysis of the H I profile sample

While I was looking to these spectra, I noticed lots of errors in the measurement of the line widths and the redshifts which were stored in catalogues. Striking examples are given in figure 2.4, but if the reader examine carefully the spectra of the figure 2.3, he will already notice some little mistakes. Several main kinds of errors seems to appear :

- A bad estimation of the **line widths** : This kind of error appears very often when there are some noise peaks close to the edges of the signal, especially when the signal level at 20% is below the noise ($0.2\,\mathbf{sl} < \mathbf{rn}$). This is the case on the first spectrum (PGC 6 173), where noise peaks are taken into account in the line width $\mathbf{W_{20}}$ on both sides of the signal.

  But this can be due also to an error in the manual data capture, which is probably what happens in the third and fourth spectra (PGC 23 781 and PGC 14 665) for $\mathbf{W_{20}}$.

- **Erroneous redshifts** : Of course, errors on the line widths have repercussions on the redshifts, as one can see on the third and fourth spectra (PGC 23 781 and PGC 14 665), where $\mathbf{V_{20}}$ is obviously wrong.

  But it appears sometimes that the redshifts were shifted with respect to the signal although the value of the width was good. This probably still reflects an error in the data capture.

- A wrong estimation of the **noise** and/or the **signal level** : On some spectra, the noise **rn** had been estimated either too low or too high. The latter case might occur when the parasites are not removed from the signal, so the noise is overestimated. However, these parasites do not appear on the spectra. This kind of mistake is illustrated on the second spectrum (PGC 23 616). **rn** is clearly shifted, as well as the signal level which is too high to be plotted on the spectrum.

These errors come from the data processing made in the radiotelescope of Nançay. Some of the measurements were done automatically and many spectra and the values of their parameters had not been checked by eye.

All this led me to do again this analysis for the 3312 spectra of the sample, by creating a program that allows to display the spectrum and redefine, by hand, the line widths, signal and noise levels. This was made with the interactive graphical routines of the PGPLOT library that provide a way to do precise measurements on a graph. All the 3312 spectra have been checked *de visu*, and in the meanwhile, a precise classification of these H I profiles was made.

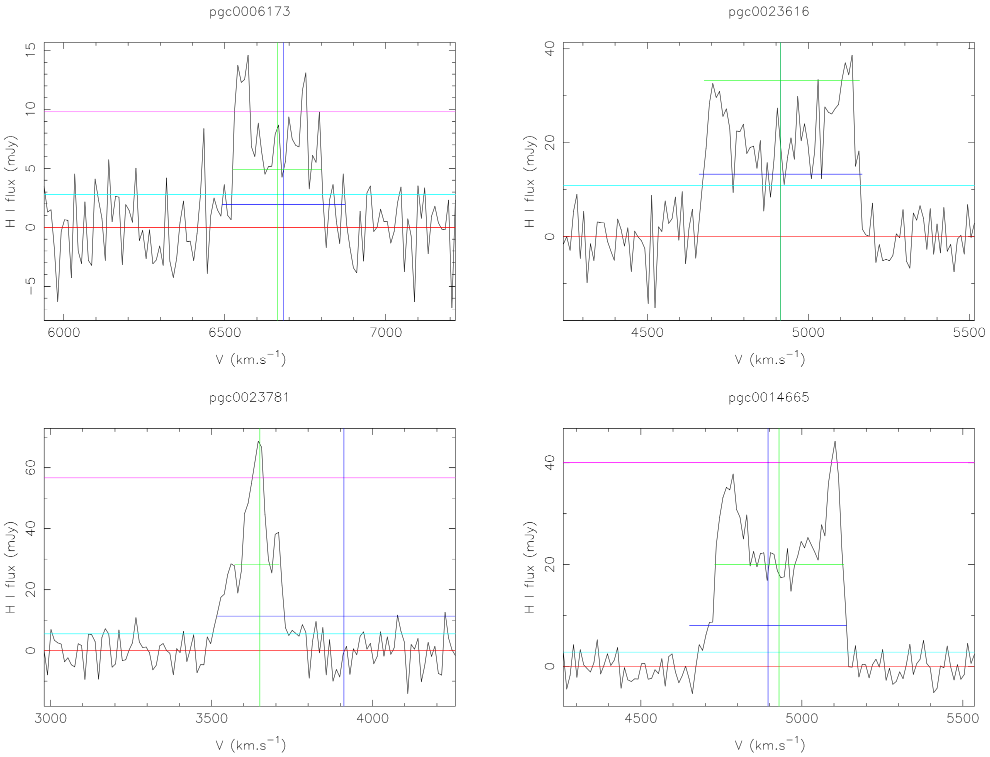


Figure 2.4: Sample of 4 H I spectra illustrating the main kinds of errors in the earlier data analysis.

2.2.3 Quality criteria and classification system

Before this study and the new analysis, a first classification was made for half of the sample. The spectra had been divided in four classes, described by the letters **A** to **D**, from the best quality data to the useless one (evocated in Paturel et al. 2003, [9]). One must emphasize the fact that such a classification is subjective and can vary from a person to another[9]. This is why, keeping the same basis, some improvements in this classification have been made. The definition of the classes is now more precise, based on new sub-classes and objective, quantitative criteria.

Firstly, one must introduce one parameter that describes roughly the quality of the signal : the so-called *signal to noise ratio* $\mathbf{sn} = \frac{\mathbf{sl}}{\mathbf{rn}}$ compares the maximum level of the profile to its noise. But this parameter is not sufficient to describe the general quality of a spectrum, in particular concerning its shape. A description of the classes, as complete as possible, and the quality criteria are presented below. For each class, a link can be done with what was said in section 2.2.1. Corresponding examples are given specifically in figure 2.5.

The **A class** concerns very intense H I lines, with high signal to noise ratio. I took $\mathbf{sn} > 9$ as the general criterion of belonging to this class. Then, this class is further divided, according to the shape of the signal :

1. **A2** spectra exhibit the typical double-horned profile, i.e. two peaks with a flat plateau between them. They are also characterized by a high degree of symmetry. Both edges of the signal must be very sharp and abrupt. The first spectrum of figure 2.5 (PGC 35 288) is an example of an **A2** profile, but also the one on figure 2.2, and the first one figure 2.3. The verticality and symmetricity can be tested by comparing the line widths at 20% and 50% levels and the corresponding recession velocities, which allows to reject the spectra that would show a possible deformation of the signal. I chose the following criteria :

   $$|\mathbf{W_{50}} - \mathbf{W_{20}}| < 25 \quad \text{and} \quad |\mathbf{V_{50}} - \mathbf{V_{20}}| < 25 \,,$$

   given that 25 km.s^{-1} corresponds approximatively to two times the resolution in the velocities.
2. The **A1** class gathers all the intense single peaks. These single horned profiles resemble a gaussian shape, with no indication of flattening. The line is not very broad, so we can test for the narrowness of the peak by fixing an upper limit for the widths. I chose the arbitrary criterion $\mathbf{W_{50}} < 100$ km.s^{-1}.

---

[9] Indeed, a bias in the classification has been noticed, depending on the optimism of the person...

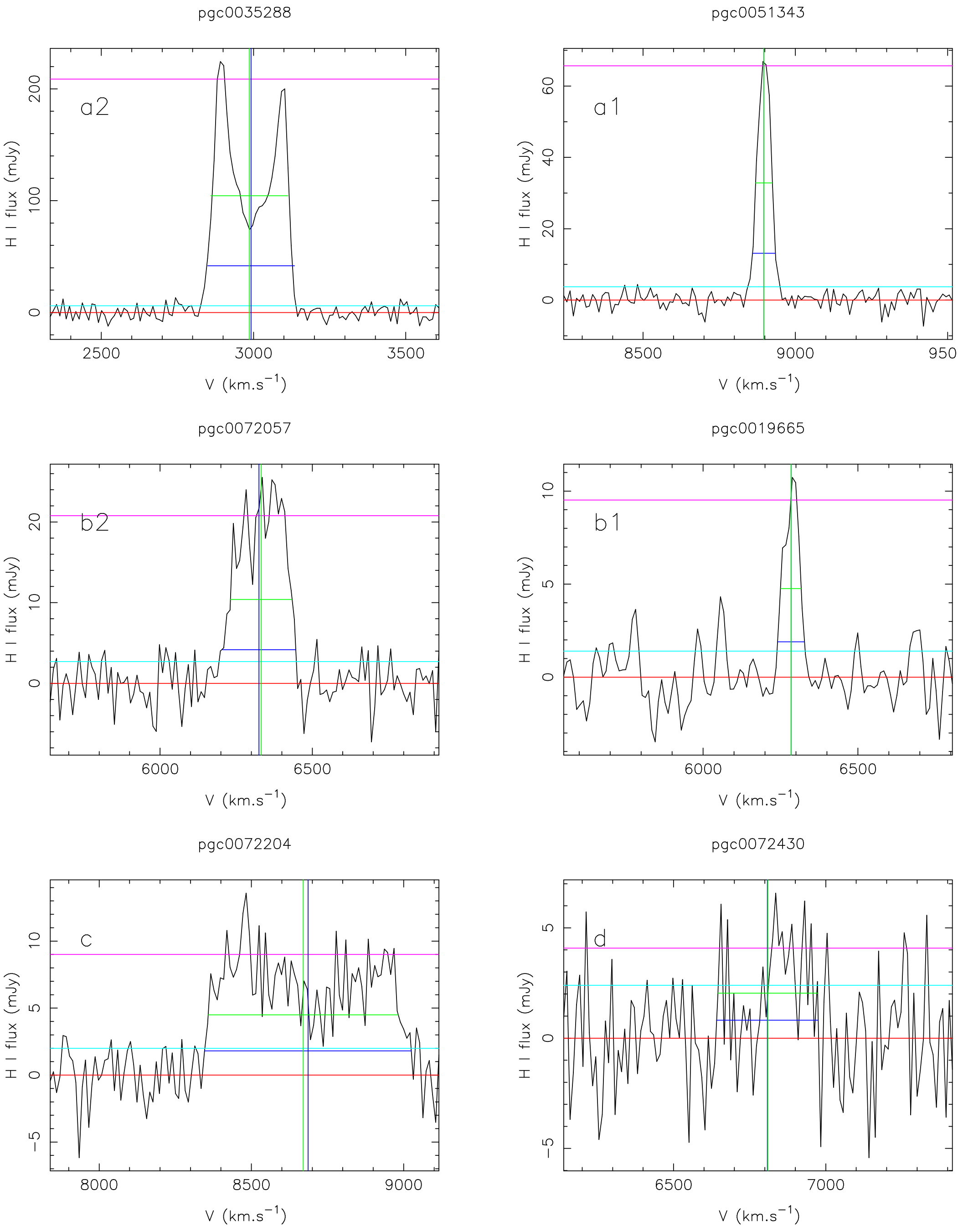


Figure 2.5: Sample of 6 H I profiles showing the different spectral classes.

B class groups together slightly more noisy spectra. In general, 90% of the **B** spectra have a signal to noise ratio $\mathbf{sn} > 5$. Many reasons can press us to classify a spectrum as a **B** instead of **A** : more noise, slight deformation of the signal, difficulty to estimate the maximum in case of an asymmetrical shape, etc. So, to find precise quantitative criteria is tricky, and these spectra have been classified "by hand", *de visu*. These spectra are sufficiently well defined so that we can still split this class in two :

1. **B2** spectra have less clear double-horned profile than the **A2**s. For instance, if we hesitate estimating the maximum of the line in case of an asymmetrical shape, or if its edges are less sharp, one will choose **B2** instead of **A2**. The third spectrum figure 2.5 (PGC 72 057) is a typical example, and one can compare the differences between **A2** and **B2**. **B2** spectra are still very reliable for TF application.
2. **B1** profiles can be more broad and more asymmetrical than the **A1**s. They show a more noisy signal, but still

with a smooth single-horned shape.

C class include the well detected but noisy profiles. The redshift measurement is still possible, but the estimation of the line widths begins to be difficult. Very few **C** spectra have a signal to noise higher than 5, but we also put in this class the very asymetrical profiles, which can have high **sn**. That is one of the reasons why these spectra are also classified by hand.

D class gathers the marginally detected and very noisy profiles, as shown on the last spectrum figure 2.5. To save time in the classification work, I used a quantitative criterion to put together these low quality signals, by comparing the 20% and 50% levels with the noise level, and testing the value of the area under the signal, so-called raw flux **rf**, expressed in Jy.km.s^{-1} and delimited by the line widths. One of the two following conditions have to be fulfilled: either the spectra is quite noisy and narrow, or, very noisy. In other words, either the 20% level is below the noise and the raw flux is small, or, the 50% level is below the noise. Specifically, one can sum up this criterion in logical terms, using the signal to noise ratio :

$$((\mathbf{sn} < 5) \text{ and } (\mathbf{rf} < 1)) \text{ or } (\mathbf{sn} < 2)$$

One of the advantages of these quantitative criteria is that they are easily programmable. These criteria have been tested while doing again the measurements on the spectra. They give good results, since 90% of the automatically classified **A2**, **A1** and **D** spectra have the required shape according to the classification that we would have made by hand. The 10% remaining are for instance very strong signals that do not fit in the criteria.

The results of the classification are given in table 2.1, where the number of spectra in each class is indicated. Note that 159 of the 559 **D** spectra are non-detected galaxies, i.e. the signal is completely flat. Thus, 3153 spectra can really be exploited, namely 3069 different galaxies, plotted figure 1.5 page 20. The total number of spectra in both classes **A2** and **B2** is 1402, which corresponds to 1377 different galaxies. These spectra represent the best data for the application of the Tully-Fisher relation.

A class		**B class**		**C class**	**D class**
A2	A1	B2	B1	1142	559
293	101	1109	108		

Table 2.1: Sharing out of the spectra within the different classes. The total number of spectra in this sample is 3312.

Note that the results of the classification in the three sub-samples REF023289, REF023359 and REF099999 are detailed in Appendix B.1.

From this classification, a sub-sample of very reliable spectra has been extracted. Now we have to take advantage of it and compare the scatter of the TF relation by selecting either the whole sample or this high quality one. The next chapter will put into practice all the notions introduced in chapter 1 (application of the Tully-Fisher relation, the normalized distance method, derivation of the unbiased distances…), and use the classification made in this one in order to do a comparative study of the results.

Chapter 3

Tully-Fisher applications: on the way to H_0…

3.1 An overview of the method

A detailed description of the examination and use of the TF relation follows, but this section starts by first explaining how we select some of the galaxies of our sample for TF studies and then, to stand back, by listing the main points of this analysis.

3.1.1 Selection of the galaxies and completeness of the sample

First of all, as introduced in section 1.2.1, the study of the TF relation requires the apparent magnitudes for as many galaxies as possible in our sample. This data is collected from the DSS[1], DENIS[2] and 2MASS[3] sky surveys, in five wavelength bandpasses: two optical, so-called B and I bands, and three near-infrared J, H and K. The associated wavelengths are given in table 3.1. These magnitudes and the measurements made on the H I spectra constitute the observational data for the following TF analysis.

In sections 1.3.2 and 1.3.3, we evoked the fact that the treatment of the second Malmquist bias, in particular by applying the normalized distance method, requires the completeness of the sample up to some magnitude limit. These completeness limits are given in table 3.1.

	B-band	I-band	J-band	H-band	K-band
λ (μm)	0.44	0.79	1.25	1.65	2.17
Surveys	DSS	DENIS	DENIS, 2MASS	2MASS	DENIS, 2MASS
Δm	$\sim$ 0.3 mag	< 0.15 mag	< 0.15 mag	< 0.15 mag	< 0.15 mag
m_{lim}	14.8	12.8	12.0	11.3	11.0

Table 3.1: The five bandpasses: corresponding wavelengths λ, sources, average uncertainty on the apparent magnitudes Δm and limiting magnitudes m_{lim} are indicated.

So, one can now sum up the main selection criteria for the further study:

- Only the complete part of the sample is retained, according to the magnitude selection. In other words, in each band, we only select galaxies with magnitudes below the limit.
- As stated in section 1.4.2, the accuracy of the apparent magnitudes must be reasonable : $\Delta m < 0.3$ appears to be a good compromise between the quality of the photometry and the fact that we need a large number of galaxies (Hanski et al. 2004, [7]).
- Finally, we exclude face-on and low inclinated galaxies, according to the discussion in section 2.1.1, so the inclination angle of the selected galaxies satisfies $i < 32°$. According to the discussion of the section 1.2.2, we also exclude galaxies having low galactic latitudes, i.e. $|b| < 15°$.

After these restrictions, we are left with a reduced but complete sample of suitable TF galaxies. The number of selected galaxies will be indicated in the table 3.2(a) of the section 3.2.2, when we will get the Tully-Fisher parameters for all the TF diagram.

---

[1] Digitized Sky Survey, `http://www-gsss.stsci.edu/DSS/dss_home-htm`

[2] DEep Near Infrared Survey of the southern sky, `http://cdsweb.u-strasbg.fr/denis.html`

[3] Two Micron All Sky Survey, `http://www.ipac.caltch.edu/2mass/`

3.1.2 From the TF relation to H_0 in short

To give an overview of the method, the main steps are listed below. Let's start by the beginning:

1. Derive the absolute magnitudes from the kinematical (redshift) distances and construct the Tully-Fisher diagrams M vs. $\log(V_{max})$ for each band.
2. Calculate the normalized distances and identify the unbiased plateau. Determine the TF relation for the unbiased galaxies.
3. Calibrate the zero-point of the TF relation in order to have the Tully-Fisher magnitudes and an absolute distance ladder.
4. Use these unbiased and calibrated Tully-Fisher parameters, i.e. the slope and zero-point of the TF relation, to estimate the true distances of the galaxies. Apply the analytical correction for the Malmquist bias, for the objects that are beyond the unbiased plateau.
5. From these distances, derive the Hubble constant H_0 for each galaxy, from the kinematical and TF absolute magnitudes. Study the distributions of H_0 values.

Each of the points summarized above are discussed in the following sections, keeping the same chronological order.

3.2 First study on the totality of the TF diagrams

3.2.1 TF diagrams for all the selected galaxies of the sample

The starting point of the TF study is to calculate the absolute magnitudes of the selected galaxies. By using the redshift z (measured on the H I spectra, cf. section 2.1.2), the apparent magnitude[4] m and assuming a value for H_0, one can derive the absolute magnitude via a simple rewriting of equation A.1, page 43 :

$$M_k = m - 5\log\left(\frac{cz}{H_0}\right) - 25\,, \tag{3.1}$$

where the index k underlines the fact that these magnitudes are calculated using the *kinematical* distances, and assuming, for instance, $H_0 = 100\,\text{km.s}^{-1}.\text{Mpc}^{-1}$.

Then, using the analysis of the H I spectra made in chapter 2, one obtains the TF diagrams for each band, plotted on figure 3.1. The linearity of the TF relation appears, but also its important scatter, as mentioned in section 1.4.2. The first advantage of the spectral classification is that one can locate on the TF diagram the nature of the spectra, by plotting each class with a different color for instance. The figure 3.1 already allows a brief qualitative analysis, and shows that the scatter is mainly due to **A1**, **B1** and **D** spectra ("warm colors"). By the way, these points make the slope of the regression less steep, and figure 3.1 clearly exhibits that the fit does not follow the dense cloud of points which corresponds to the more reliable data, i.e. **A2**, **B2** and **C**, in "cold" colors. The scattering caused by **D** spectra is mostly due to their poor quality and the big uncertainty on the linewidths and redshift measurements.

But why **A1** and **B1** do not follow the TF relation ? Because of the selection, these points cannot be face-on galaxies (cf. section 2.2.1 on the interpretation of these kind of profile). If they are dwarf galaxies, they are not expected to follow the TF relation, so they should be removed from the sample. Finally, we thought that they are mostly confused galaxies, as evoked in section 2.2.1.

One can also note the differences between the bands. TF parameters depend on the wavelength of the bandpass: for instance, I-band and near-infrared ones show less scattering than the B-band, essentially because near-infrared luminosities trace better the mass than the B-band. Moreover, they also suffer less from absorption or inclination corrections.

3.2.2 TF parameters: slope, zero point and scatter

In order to get the TF parameters, we fit the data thanks to the least squares method. More precisely, we want to fit a set of N data points (M_i, p_i) to a straight-line model $M = a_1 p + a_0$, where $p = \log(V_{max})$ is the rotational parameter of the TF relation. Assuming that all the individual uncertainties on M_i are equal, the slope a_1 and zero-point a_0 are determined by minimizing the *chi-square* function, which in the case is:

$$\chi^2(a_1, a_0) = \sum_{i=0}^{N} (M_i - a_1 p_i - a_0)^2\,. \tag{3.2}$$

---

[4] Note that the apparent magnitudes are also corrected for inclination, and extinction due not only to our Galaxy but also to the internal absorption of the observed galaxy.

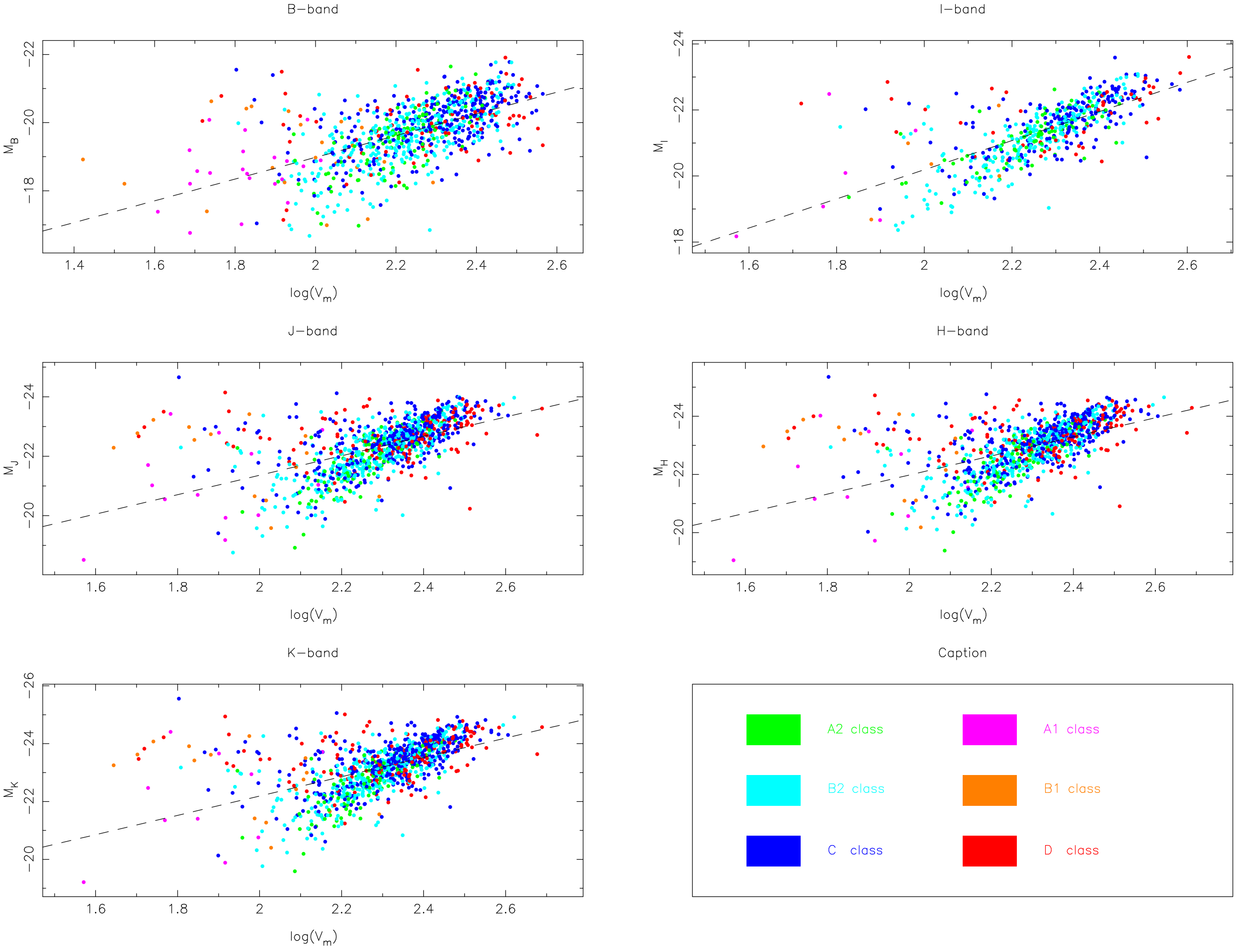


Figure 3.1: Tully-Fisher diagrams in the five bands B, I, J, H and K for all the selected galaxies, and all the classes. Note that the more negative the magnitude is, the more luminous is the galaxy. The dashed line is the least squares fit on all the points.

Then the scatter of the TF relation is characterized by the quantity:

$$\sigma_{TF} = \sqrt{\frac{\chi^2}{N-1}}, \tag{3.3}$$

where χ^2 is the effective minimum of the equation 3.2.

Even if the TF parameters we are interested in are those of the unbiased part of the diagram, some preliminary values for the whole TF diagram are given in table 3.2, in order to further quantify and clarify the first results mentioned above, in section 3.2.1.

Firstly, the table 3.2(a) shows that the selection made for the TF study reduces a lot our sample, with respect to the 3 069 galaxies available at the begining. This is mostly due to the fact that we do not have all the apparent magnitudes of all our galaxies, but only for half of them in I-band for instance. Then, on the same table, the comparison of the scatter values suggests that the I band has the lowest scatter, despite the small number of points, which is partly due to its accuracy.

Table 3.2(b) gives the characteristics of the TF relation when we only select **A2** galaxies. Even if we reduce the number of points by selecting high quality data, we still reduce the scatter, by a factor 20% to 30% according to the bands. A more precise comparative study will be done for the unbiased galaxies, in section 3.3.2.

Finally, the table 3.2(c) confirms what was stated before about **A1** and **B1** spectra: these galaxies do not follow any TF relation and have a large scatter, so that the fit does not have a lot of sense here. By comparing tables 3.2(a) and 3.2(b), one sees again that these galaxies are responsible for flattening the slope of the fit made on all the classes (cf. also figure 3.1). Definitely, we must exclude these galaxies for a precise application of the TF relation.

	a_1	a_0	σ_{TF}	N
B	-3.19	-12.60	0.74	842
I	-4.41	-11.37	0.68	535
J	-3.27	-14.82	0.76	973
H	-3.28	-15.42	0.76	977
K	-3.34	-15.51	0.77	986

(a) All the selected galaxies and all the classes. The number of points, slope and zero-point values corresponds to the TF diagrams of the figure 3.1.

	a_1	a_0	σ_{TF}	N
B	-5.68	-6.93	0.64	106
I	-4.73	-10.54	0.49	88
J	-6.41	-7.39	0.59	105
H	-6.72	-7.25	0.60	107
K	-6.78	-7.32	0.60	104

(b) Results of the fit only made on the **A2** galaxies.

	a_1	a_0	σ_{TF}	N
B	-0.06	-18.93	0.97	37
I	-3.89	-12.76	1.35	11
J	-1.06	-19.28	1.45	20
H	-0.30	-21.50	1.35	33
K	-0.29	-21.73	1.37	33

(c) Results of the fit only made on the **A1** and **B1** galaxies.

Table 3.2: Tully-Fisher parameters for the whole diagrams in each band: slope, zero-point, scatter and number of points.

3.3 Investigation of the unbiased part

This section aims at presenting, in a simple manner, how we calculate the TF relation for the unbiased part of the TF diagram, using the method described in chapter 1.

3.3.1 The normalized distance method in practice

The second Malmquist bias is the consequence of the cut-off in the distribution of magnitudes, as discussed in section 1.3.2. The normalization of distances, explained in section 1.3.3, may be re-formulated by considering two quantities:

- The *normalized distance modulus*, corresponding to the normalized distance d_n given in section 1.3.3, is defined as:

  $$X_n = \mu_k + M_{TF} - m_{lim} = M_{TF} - M_{lim} \, , \tag{3.4}$$

  where μ_k is the kinematical distance modulus, M_{TF} the absolute magnitude derived from the Tully-Fisher relation, m_{lim} and M_{lim} the apparent and absolute limiting magnitudes. X_n realises the same change of scale for the distance modulus than d_n for the distances. This normalization amounts to a "shift" in the kinematical distance moduli so that the amount of Malmquist bias is the same for all the galaxies having the same X_n.

- The *normalized magnitude*

  $$Y_n = M_{TF} - M_k \, , \tag{3.5}$$

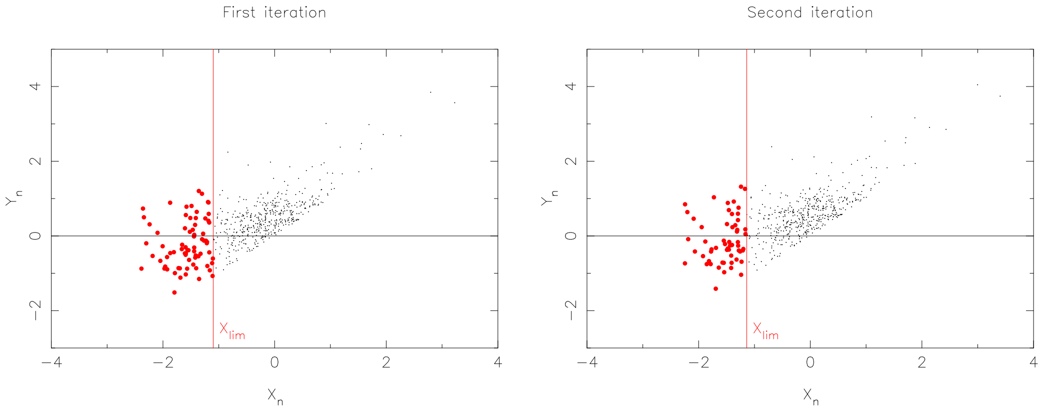
  where $M_k = m - \mu_k$, measures the deviation of a point from the fitted TF relation, so Y_n is also more precisely called *TF residual*.

The Y_n vs. X_n plot is called normalized distance diagram, and allows to determine the unbiased plateau. How exactly ? And how do we get the unbiased TF parameters ? An iterative method is used and explained below. Its illustration on our data is given in figure 3.2.

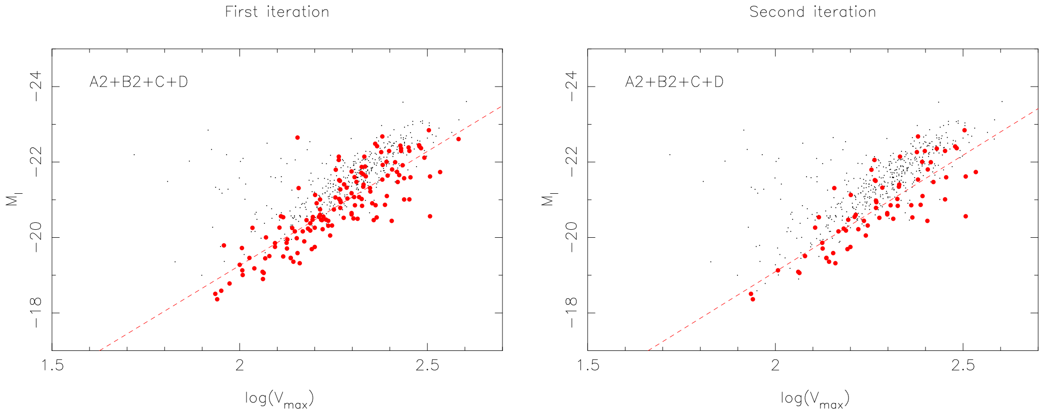
1. In the first iteration, one calculates X_n and Y_n from M_k (equation 3.1) and M_{TF}, using some initial values of the TF parameters. These input parameters are given in Appendix B.2. We determine the limit X_{lim} from which the bias appears by using the scatter of the previous TF relation, assuming that the distribution of residuals is gaussian, i.e. $X_{lim} = -2\sigma_{TF}$. Then, one selects the unbiased galaxies having $X_n < X_{lim}$ (red points on the figure 3.2) and determine with them a first set of unbiased TF parameters (a_1, a_0, σ_{TF}).
2. Then, one does again the same loop "TF-parameters $\Rightarrow$ normalized-distance $\Rightarrow$ unbiased-subsample $\Rightarrow$ TF-parameters" but starting with the unbiased TF parameters derived at the end of the first iteration. So we have now our final TF parameters.

This iterative method provides a good accuracy for the determination of the unbiased plateau. The normalized distance diagrams of the figure 3.2(a) clearly show the magnitude cut-off in the distribution of TF residuals, meaning that all the points at $Y_n < X_n$ are rejected, because of the completeness limit. The unbiased plateau ($X_n < X_{lim}$) is the domain where the average of Y_n stays close to 0. Beyond the plateau, the bias is clearly visible.

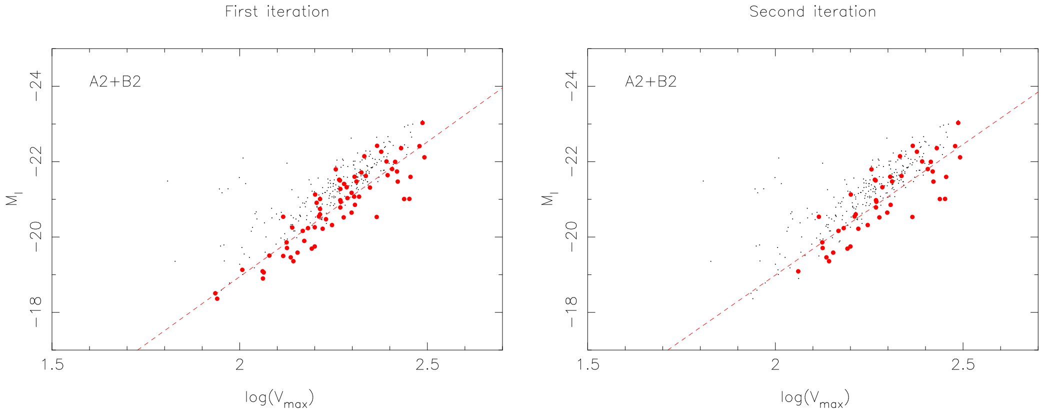
The TF diagrams on figures 3.2(b) and 3.2(c) underline the distribution of the unbiased galaxies in the TF diagram. One can now quantify the TF parameters derived with this method.



(a) TF residuals for the two iterations, with (**A2**+**B2**+**C**+**D**) spectral classes. The unbiased plateau is the $X_n < X_{lim}$ region. These galaxies are then selected for the determination of the unbiased TF relation (cf. figure 3.2(b)). One clearly sees the completeness limit, corresponding to the sharp edge $Y_n = X_n$ in the distribution of the TF residuals.



(b) TF diagrams for the two iterations, with (**A2**+**B2**+**C**+**D**) spectral classes. The red dashed line fits the unbiased galaxies (red points), and so represents the unbiased TF relation.



(c) TF diagrams for (**A2**+**B2**) spectral classes, for a comparison with figure 3.2(b).

Figure 3.2: Normalized distance and TF diagrams for I band. Red points are the galaxies that belong to the unbiased plateau. All the small black dots are galaxies beyond the plateau, so affected by the selection bias.

3.3.2 Comparative study of TF parameters for the unbiased galaxies

One can already notice, qualitatively, the influence on the selection of good quality data by comparing figures 3.2(b) and 3.2(c). The scattering of the biased and unbiased galaxies is clearly reduced and the accuracy on the TF parameters improved.

The unbiased TF parameters derived after the second iteration are given in table 3.3. These five tables show the values obtained for different combinations of spectral classes.

	a_1	a_0	σ_{TF}	N
B	-10.08	3.70	0.60	22
I	-6.38	-6.42	0.49	16
J	-7.67	-4.31	0.56	20
H	-7.56	-5.21	0.54	19
K	-5.63	-10.03	0.51	16

(a) **A2** galaxies.

	a_1	a_0	σ_{TF}	N
B	-7.17	-2.90	0.67	61
I	-6.94	-5.10	0.54	45
J	-7.62	-4.28	0.58	79
H	-7.80	-4.46	0.58	76
K	-8.11	-3.94	0.57	75

(b) **A2** and **B2** galaxies.

	a_1	a_0	σ_{TF}	N
B	-6.45	-4.43	0.66	91
I	-6.63	-5.75	0.57	70
J	-7.49	-4.58	0.58	161
H	-7.61	-4.94	0.58	159
K	-7.90	-4.48	0.58	155

(c) **A2**, **B2** and **C** galaxies.

	a_1	a_0	σ_{TF}	N
B	-5.28	-7.07	0.69	100
I	-6.18	-6.74	0.60	70
J	-6.31	-7.21	0.65	146
H	-6.49	-7.42	0.65	151
K	-7.09	-6.30	0.61	162

(d) **A2**, **B2**, **C** and **D** galaxies.

	a_1	a_0	σ_{TF}	N
B	-5.15	-7.32	0.67	100
I	-6.35	-6.33	0.58	72
J	-5.88	-8.18	0.67	161
H	-5.88	-8.81	0.67	162
K	-6.55	-7.55	0.64	173

(e) All classes of spectra.

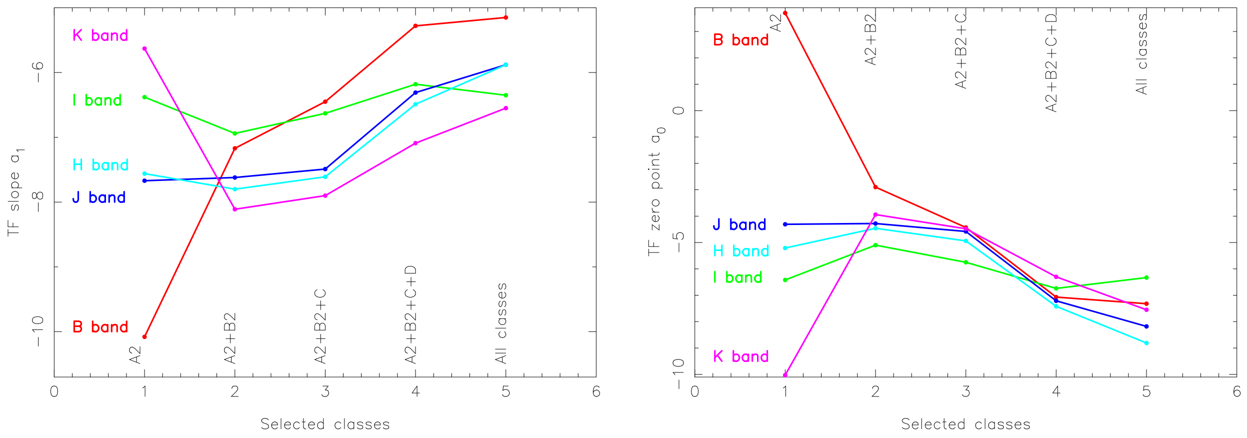
Table 3.3: Unbiased TF parameters in each band: slope, zero-point, scatter and number of points in the unbiased plateau after the second loop.

Notice that the percentage of points in the unbiased region, with respect to all the selected galaxies, increases if one only selects good quality data. When one selects all the spectral classes (table 3.3(e)), this percentage is about 10%, as in the previous KLUN studies.

One can see the influence of the unbiased domain by comparing tables 3.2(a) and 3.3(e). The important reduction of the number of galaxies in the unbiased plateau means that we need a large sample of galaxies. In particular, in table 3.3(a), where we only have the **A2** galaxies, the unbiased population is so small that we cannot fully trust the derived unbiased TF parameters. Indeed, there we see some big variations in the slope and zero-points between the bands. However, the TF scatter is small, so there are few points, but well distributed. Anyway, these values have to be taken cautiously.

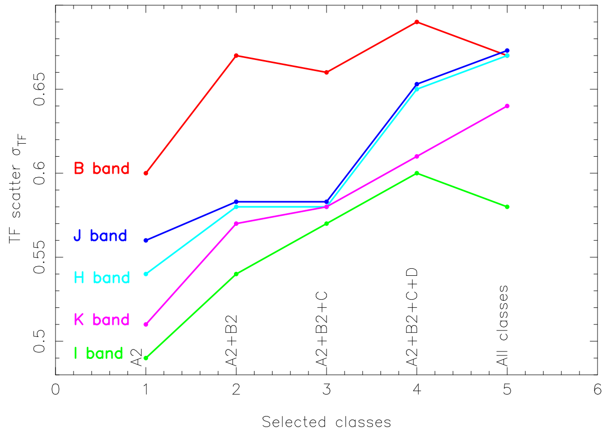
The results of the table 3.3 concerning the slope, zero-point and the scatter are plotted in figure 3.3, which provides a comparative view of the TF parameters between the different sets of spectral classes, but also between the different bands. The main results are the following:

- At a given set of spectral classes, one can compare the evolution of the TF parameters between the different bandpasses, i.e. with the wavelength. The figure 3.3(a) shows a general trend of steepening TF slopes towards longer wavelengths. Briefly, the physical reason for that behaviour is that we do not detect the same types of stars when we look through different bandpasses, so we do not exactly trace the same distribution of mass in the galaxy.
- On figure 3.3(c), one sees a general increase of the TF scatter, caused by an increasing number of low quality data. However, as mentioned above, selecting only **A2** galaxies do not lead to very accurate TF parameters, so one has to find the best compromise between a small TF scatter and a statistically robust sample. By comparing figures 3.3(a), 3.3(b) and 3.3(c), it appears that the best compromise is realised by the (**A2**+**B2**) and (**A2**+**B2**+**C**) sets of galaxies. Moreover, one finds again that the I-band is the most accurate bandpass, leading to the smallest TF scatter. This is essentially due to the very good quality of the photometric measurements for this band.
- When one only selects **A2** galaxies, the scattering of the parameters between the bands is large and random, indicating again the lack of accuracy of this set because of its small population in the unbiased plateau.
- Finally, comparing figures 3.3(a) and 3.3(c), note that a larger scatter always causes a flatter slope: a bigger uncertainty on $\log(V_{max})$ makes the slope shallow on the TF diagram.



(a) Evolution of the slope a_1 with respect to the selected classes of spectra.

(b) Evolution of the zero-point a_0 with respect to the selected classes of spectra.



(c) Evolution of the scatter σ_{TF} with respect to the selected classes of spectra.

Figure 3.3: Influence of the quality of the spectra on the unbiased TF slope, zero-point and scatter. X-axis indicates which spectral classes are included. The population of the sample increase from left to right.

Now that we have the Tully-Fisher parameters in the unbiased plateau, we can combine these results with the data of calibrators in order to derive the Hubble constant. Again, the aim of this study is to compare the results and the scatter of the distribution of H_0 between the different sets of spectral classes we select.

3.4 Towards an assessment of the Hubble constant

3.4.1 Calibration and correction for the Malmquist bias

For the kinematical distances used for the TF relations above, an arbitrary value of Hubble constant was used: $H_0 = 100\,\mathrm{km.s^{-1}.Mpc^{-1}}$. Now, in order to calculate the absolute zero-point of the TF relation, and consequently the absolute distance ladder, one must use a sample of calibrators. I do not discuss here in details the choice of this sample, but most of them are primary distance indicators, i.e. their absolute distance is known *independently of the Hubble constant*, thanks to the detection of Cepheid stars in these galaxies (cf. section 1.1.1 and appendix A).

The principle of the calibration is quite simple: keeping the unbiased slope a_1, determined after the second loop of plateau determination, we adjust the zero point on the distribution of the calibrators in the TF diagram. The new a_0 is then the average of the calibrators's zero points, calculated with a_1:

$$a_0 = \frac{\sum_{i=1}^{N_{cal}} \left(M_i^{cal} - a_1 \log V_{m_i}^{cal} \right)}{N_{cal}} \,, \tag{3.6}$$

where N_{cal} is the number of calibrators, M_i^{cal} and $V_{m_i}^{cal}$ their absolute magnitudes and rotation speeds. The calibration method is illustrated on the figure 3.4.

Calibrators, the unbiased TF relation, determined after the second loop of plateau determination, and the calibrated one, are plotted on the TF diagrams in each band. After the calibration, the zero point of the unbiased TF relation is shifted, as well as the normalized magnitude Y_n.

Then, we apply the correction for the second Malmquist bias to the galaxies that are beyond the plateau. We remind the result evoked in section 1.3.3 by rewriting equation 1.7. The corrected normalized magnitudes Y_n^{cor} are thus given by:

$$Y_n^{cor} = Y_n + f(X_n, \sigma_{TF}, m_{lim}) \,. \tag{3.7}$$

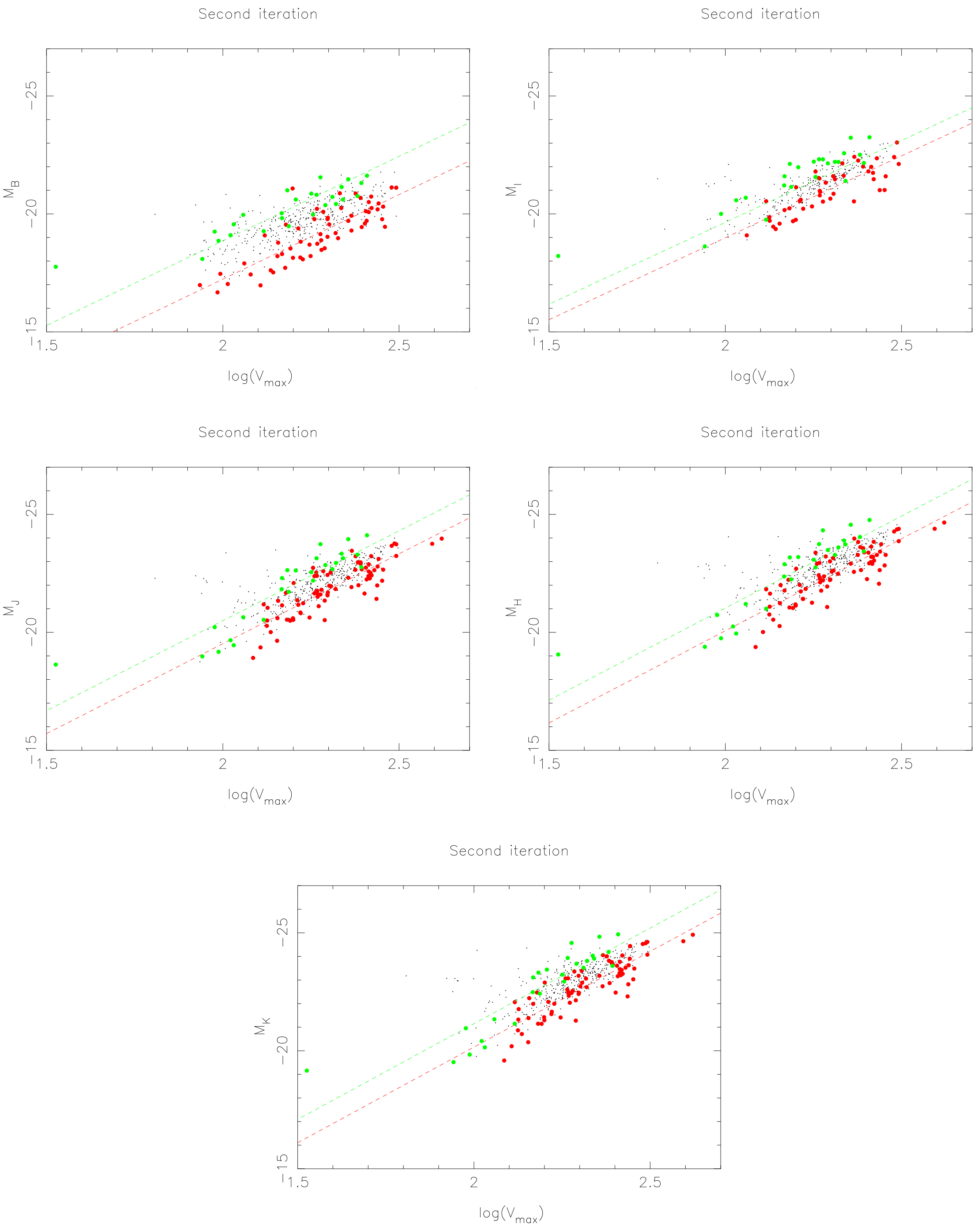


Figure 3.4: Illustration of the calibration: Tully-Fisher diagrams, unbiased and calibrated TF relations for the five bands. Red and green points are the unbiased and calibrator galaxies. The previous unbiased TF relation (red dashed line) is "translated" on the calibrators: the green dashed line is then the final, calibrated TF relation. Note that only **A2** and **B2** spectra have been selected here.

3.4.2 Distributions of H_0

We have now the true, Malmquist corrected, distances of all the selected galaxies. From Y_n^{cor}, it is quite straightforward to derive H_0 for each galaxy. According to the definitions of the normalized magnitude (equation 3.5), the kinematical distance modulus (equation A.1), and given that we have assumed $H_0 = 100\,\mathrm{km.s^{-1}.Mpc^{-1}}$ to derive the latter, one obtains:

$$\log H_0 = 0.2\,Y_n^{cor} + 2\,. \tag{3.8}$$

The distributions of H_0 are plotted on the histograms of figure 3.5. Here is the way we analyse these distributions: First, for each band, we calculate the average $\langle H_0 \rangle$ and the standard deviation of the distribution σ_{H_0}, characterizing its width, and calculated as follows:

$$\sigma_{H_0} = \sqrt{\frac{\sum_{i}^{N_{val}} \left(H_{0_i} - \langle H_0 \rangle\right)^2}{N_{val}}}\,, \tag{3.9}$$

where N_{val} is the number of H_{0_i} values in the considered distribution.

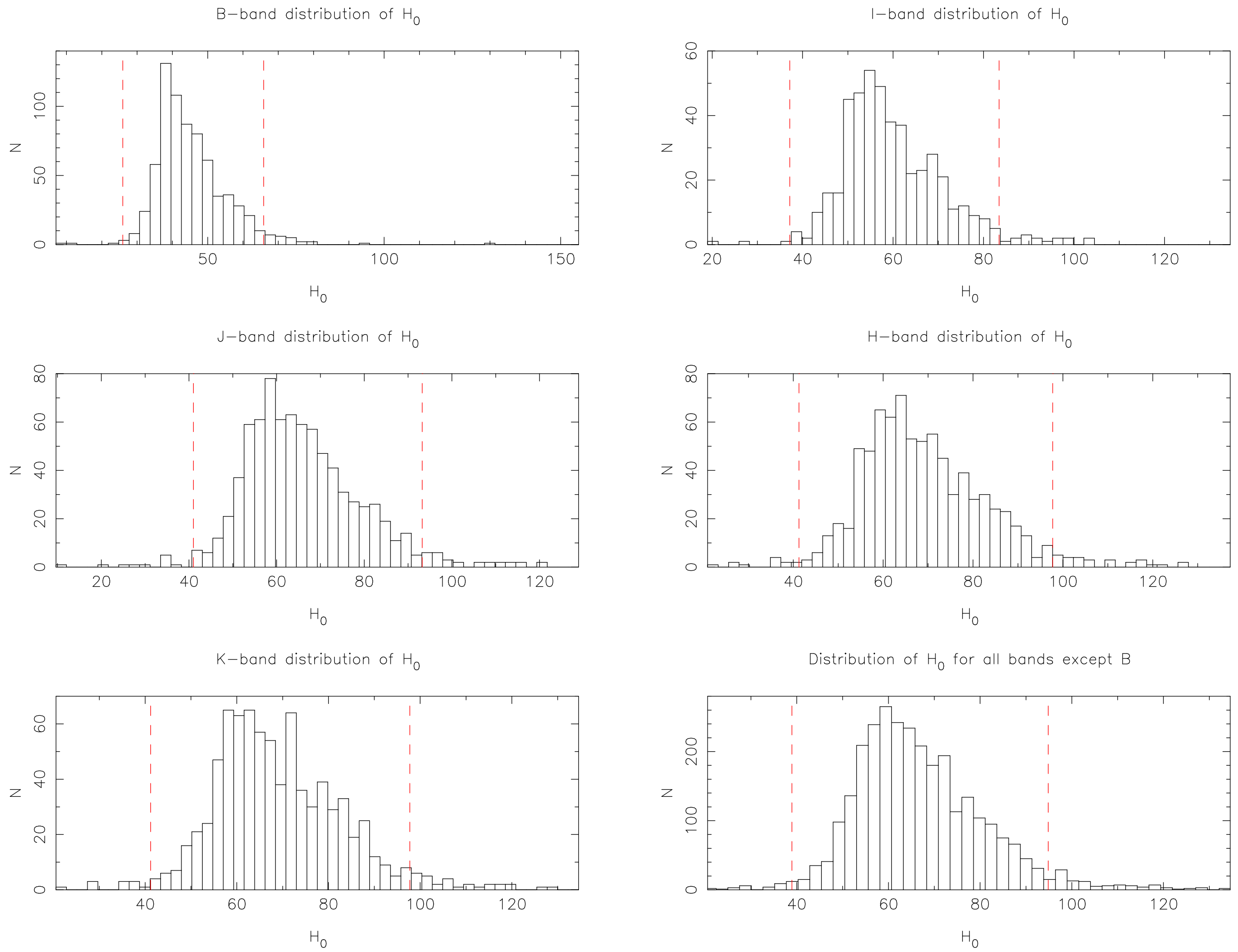


Figure 3.5: Histograms of the distributions of H_0 for the five bandpasses and for the (I+J+H+K) set of bands. The X-axis range is divided into 50 equal bins. N is the number of H_0 values in each bin. For these histograms, the set of (**A2**+**B2**+**C**) was chosen. Red dashed lines are the $\pm 2\sigma_{H_0}$ limits from the mean value $\langle H_0 \rangle$.

Then we apply a $2\sigma_{H_0}$ rejection, i.e. we eliminate all the discrepant values out of the range $[\langle H_0 \rangle - 2\sigma_{H_0}; \langle H_0 \rangle + 2\sigma_{H_0}]$. The two boundaries of this range are plotted on the figure 3.5. After that, we calculate again the mean, which corresponds to the adopted value of the Hubble constant for the considered band, and the value of the sigma. And finally, we do the same analysis for the whole set of H_0 values in the four bands I, J, H and K, excluding B band because its lack of accuracy and the discrepancy of the values of H_0 obtained with this band, compared to the others. This disagreement between the bands will be discussed later.

3.4.3 Comparative study of the values of H_0 and σ_{H_0}

All the values of $\langle H_0 \rangle$ and σ_{H_0} are indicated in the table 3.4. The KLUN team has previously obtained $H_0 \simeq 55$ km.s^{-1}.Mpc^{-1} using the unbiased TF relation, but only with B band, and without any selection of the quality of the H I spectra. In order to compare easily the results, I summarize them in the figure 3.6, which is the same kind of graph than the figure 3.3.

bands	$\langle H_0 \rangle$	σ_{H_0}
B	44.7	9.3
I	66.2	9.7
J	70.3	13.3
H	73.5	13.7
K	84.3	16.1
“All”	73.1	14.5

(a) **A2** galaxies.

$\langle H_0 \rangle$	σ_{H_0}
45.9	8.4
60.5	8.8
65.6	10.5
68.3	11.2
69.6	11.6
66.5	11.2

(b) **A2** and **B2** galaxies.

$\langle H_0 \rangle$	σ_{H_0}
44.3	7.9
59.1	9.3
65.1	10.9
68.1	11.5
68.1	11.5
65.7	11.3

(c) **A2**, **B2** and **C** galaxies.

$\langle H_0 \rangle$	σ_{H_0}
45.5	8.4
57.5	9.8
63.7	12.0
65.9	12.6
67.6	12.7
64.4	12.5

(d) **A2**, **B2**, **C** and **D** galaxies.

$\langle H_0 \rangle$	σ_{H_0}	bands
44.7	8.4	B
57.5	9.9	I
60.8	11.8	J
62.9	12.6	H
65.5	13.0	K
62.2	12.4	“All”

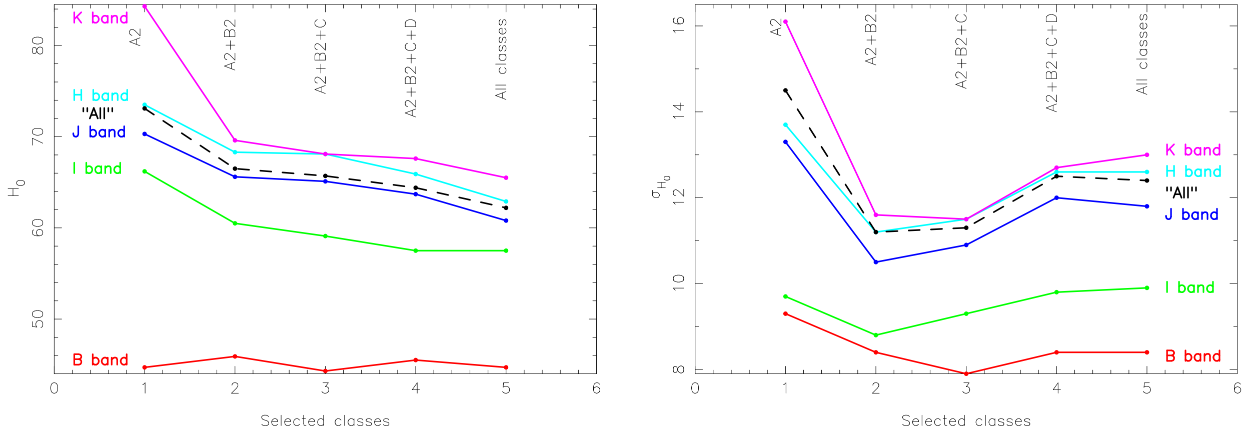
(e) All classes of spectra.

Table 3.4: Hubble constant mean values and standard deviation of its distribution in each bandpass. The “All” row corresponds to the compilation of the bands I+J+H+K.

Firstly, the values of H_0 show some big differences between the bands. Particularly, as noted above, the B-band provides low values of H_0 compared to the other bandpasses. The reason for the discrepant values of H_0 obtained with B, I, or 2MASS magnitudes is not completely clear. What is sure is that these differences come from the distribution of the calibrators. Indeed, a closer look at them, cf. figure 3.4, shows that the TF slope of the calibrator sample is steeper than for the plateau galaxies. This feature is more pronounced for the 2MASS wavelengths, and might indicate some systematic trend between the different surveys. This remains to be confirmed.

Coming from the same 2MASS survey, J, H and K bands are in good agreement, and give very similar values for H_0. Because of the relative inaccuracy of the B-band and the fact that its derived values of H_0 are quite inconstitent with the other bands, one decides not to take it into account when one gathers the data of all the bands to study the complete distribution (last histogram of the figure 3.5).

Figure 3.6(b) shows that the choice of the (**A2**+**B2**), or possibly (**A2**+**B2**+**C**), minimizes the scattering of the values between the bands and the scatter of the H_0 distributions themselves.



(a) Evolution of the Hubble constant H_0 with respect to the selected spectral classes.

(b) Evolution of the scattering of the H_0 distribution with respect to the selected spectral classes.

Figure 3.6: Influence of the quality of the spectra on the Hubble constant and the scatter of its distribution. The X-axis of the plots indicates which classes of galaxies are included. The population of the sample increase from left to right.

3.4.4 Uncertainties on the Hubble constant itself

As stated above, the standard deviation σ_{H_0} of the distribution of the Hubble constant only reflects its width. One calculates the uncertainty on the mean value $\langle H_0 \rangle$ from the distribution of the $\log H_0$ values. The uncertainty on the log of the mean value $\log\langle H_0 \rangle$ contains four contributions:

$$\Delta \log\langle H_0 \rangle = \sqrt{\sigma^2_{\log\langle H_0 \rangle} + \sigma^2_{\mu_{cal}} + \sigma^2_{m_{cal}} + \sigma^2_{TF_{cal}}}\,, \qquad (3.10)$$

where $\sigma_{\log\langle H_0 \rangle}$ is the uncertainty on the mean value of the $\log H_0$ distribution itself, $\sigma_{\mu_{cal}}$ the error on the distance modulus of the calibrators calculated according to the value of Paturel et al. 2002 [10], $\sigma_{m_{cal}}$ the uncertainty on the apparent magnitude of the calibrators, and $\sigma_{TF_{cal}}$ is related to the scatter of the TF relation for the calibrators, and is actually the error on the zero point.

The results are given in table 3.5. With this study, combining the data of the four infrared bands, we provide the following value of the Hubble constant:

$$H_0 = 66 \pm 11\,\text{km.s}^{-1}.\text{Mpc}^{-1}$$

This value of H_0 is in agreement with many previous studies: The Hubble Space telescope team finds $H_0 = 72 \pm 8$ (Freedman et al. 2001), the KLUN team $H_0 = 55 \pm 7$ (Theureau et al. 1997b [5]), Giovanelli et al. 1997 finds, in I-band, $H_0 = 69 \pm 5$, etc. Our uncertainty on H_0 is in good agreement with the study of Paturel et al. 2002 [10]. They conclude that when one takes into account all the uncertainties on the calibrators, the error on their distance modulus can barely be better than $\sigma_{\mu_{cal}} = 0.3$ mag, and so the uncertainty on H_0 cannot be better than 10 km.s^{-1}.Mpc^{-1}.

bands	$\langle H_0 \rangle$	ΔH_0	$\langle H_0 \rangle$	ΔH_0	$\langle H_0 \rangle$	ΔH_0	$\langle H_0 \rangle$	ΔH_0	$\langle H_0 \rangle$	ΔH_0	bands
B	44.7	8.5	45.9	7.7	44.3	7.3	45.5	7.4	44.7	7.2	B
I	66.2	10.8	60.5	9.8	59.1	9.7	57.5	9.4	57.5	9.4	I
J	70.3	11.8	65.6	11.1	65.1	10.9	63.7	10.9	60.8	10.5	J
H	73.5	12.4	68.3	11.5	68.1	11.5	65.9	11.3	62.9	10.9	H
K	84.3	14.7	69.6	11.8	68.1	11.5	67.6	11.5	65.5	11.2	K
"All"	73.1	14.0	66.5	11.5	65.7	11.4	64.4	11.3	62.2	11.1	"All"

(a) **A2** galaxies.

(b) **A2** and **B2** galaxies.

(c) **A2**, **B2** and **C** galaxies.

(d) **A2**, **B2**, **C** and **D** galaxies.

(e) All classes of spectra.

Table 3.5: Hubble constant mean values and their uncertainties in each bandpass. The "All" row corresponds to the distribution of all the bands, except the B one, because of its discrepancy.

Conclusion

In the general investigation of the Tully-Fisher relation, one of the most popular extragalactic distance indicators, improvements have been made concerning the classification of the H I spectra. Based on some objective and quantitative criteria, the profiles are categorized according to their quality and shape and their distribution in the Tully-Fisher diagram is studied. In particular, very low quality data is found to be an important source of scatter in the TF relation. Moreover, smooth single horned spectra (classes **A1** and **B1**) are "outsiders" in the TF relation, and we suspect that this is due to some confusions in the detection.

Interesting results have been obtained concerning the behaviour of the TF parameters according to the set of spectral classes and the bandpass we select:

- Firstly, generally speaking, the I-band is the one that shows the smallest scatter, despite a lower number of points compared to the other bands. This is explained by the quality of the photometry for this bandpass.
- There is a general trend for all the five used bands in increasing the TF scatter when one adds poor quality spectra in the sample, meaning that a high number of points do not necessarily compensate the scatter. A compromise has to be found between quality and quantity. It appears that the (**A2**+**B2**) or (**A2**+**B2**+**C**) sets of classes lead to the best equilibrium. Indeed, for these collections of spectra, the scattering of the TF parameters is the lowest.

Of course, these results have repercussions on the distributions of the Hubble constant, and the accuracy on the mean is the best for the selection of classes mentioned above. Besides, the values given by the four infrared bands are in very good concordance with each other. Combining the data for these bands and selecting the good quality sets of spectra, we find:

$$H_0 = 66 \pm 11\,\mathrm{km.s^{-1}.Mpc^{-1}}$$

On the contrary, the B-band leads to small values of $H_0 \simeq 45$. This is caused by some differences in the calibrators' magnitudes between the different photometric surveys. It remains to be seen whether this trend is systematic or not.

Of course, this study is more a start than an end in itself ! Indeed, this is the beginning of a precise classification an selection of the H I spectra for precise Tully-Fisher applications. For instance, this classification has to be done on a larger sample of galaxies, given that more than 16 000 TF galaxies are now available. This would certainly lead to more robust TF parameters, and thus less uncertainties on galaxy distances and on the Hubble constant.

In the meanwhile, a very interesting study could concern the link between the belonging to one class of spectra and other physical characteristics of the galaxy in question: its morphology, brightness profile, environment, etc. This could lead to further refinements of the TF relation.

Finally, I thank again all my supervisors, for having really involved me in their research project. Besides, this work will lead to a publication, and the creation of a new database of H I profiles, including the new analysis and the spectral classification done during this study.

This training brings me all the "feelings" that the research work can bring: happiness, satisfaction, but also moments of disappointment and irritation when one fights with the data ! But anyway, one feels so excited when the value of the Hubble constant appears on the screen…

Appendix A

Glossary

- parallax: The trigonometric parallax (π) of a star is the angle subtended by 1 astronomical unit[1] at the star's distance (d) from the Sun.

  For convenience, the astronomer's unit of distance, the parsec[2] (pc), is defined as the distance at which a star's parallax would be one second of arc (arcsec), the unit in which parallaxes are traditionally measured. This gives the simple relation:

  $$d = \frac{1}{\pi}$$

  The measurement of parallaxes is the most basic tool to estimate stellar distances in our Galaxy, and is the first step in establishing the cosmic distance scale.

- Cepheid variables and period-luminosity relation: A Cepheid variable is a star whose light output, surface temperature, and spectrum change because of a periodic expansion and contraction of the star's upper layers. They are also called pulsating stars.

  The period of pulsation is directly related to the luminosity of the Cepheid: the longer the period, the greater the mean intrinsic brightness. This correlation[3] makes Cepheids important cosmic distance indicators, since, by measuring a Cepheid's period, deriving its intrinsic brightness, and then comparing it to the observed one, the star's distance can be worked out.

- apparent and absolute magnitudes: The apparent magnitude m of an object is a measure of the intensity of radiation, within a particular wavelength interval, received from that object at the Earth. The absolute magnitude M is the magnitude that the object would have if it was situated at a standard distance of 10 parsecs, and is directly related to the intrinsic luminosity of the object in question.

  Magnitudes are expressed in an inverse logarithmic scale, meaning that the more negative the magnitude, the brighter the object.

- distance modulus: The distance modulus μ is the difference between the apparent and the absolute magnitudes. Moreover, this quantity is related to the distance of the object:

  $$\mu = m - M = 5\log(d) + 25 \quad \text{where} \quad d = \frac{c\,z}{H_0} \quad \text{is in Megaparsecs (Mpc).} \tag{A.1}$$

---

[1] The astronomical unit is the distance from Earth to the Sun. $1\,AU \simeq 150.10^6$ km.

[2] 1 pc $\simeq 3.26$ light-year

[3] discovered by Henrietta Leavitt in 1912.

Appendix B

Additional tables

B.1 Detailed results of the spectral classification for the 3 sub-samples

A class		B class		C class	D class
A2	A1	B2	B1		
201	44	634	51	642	182

(a) REF023289 sub-sample (Theureau et al. 1998, [6]). The total number of spectra is 1754.

A class		B class		C class	D class
A2	A1	B2	B1		
68	37	302	38	255	102

(b) REF023359 sub-sample (Paturel et al. 2003, [9]). The total number of spectra is 802.

A class		B class		C class	D class
A2	A1	B2	B1		
24	20	173	19	245	275

(c) REF099999 sub-sample (Theureau et al. 2004). The total number of spectra is 756, but 159 of the 275 **D** spectra are undetected galaxies, so we are left with 597 profiles where a signal is effectively detected.

Table B.1: Sharing out of the spectra within the different classes for the three sub-samples. The total number of spectra is 3312.

B.2 Starting values of the TF parameters for the iterative normalized distance method

	a_1	a_0	σ_{TF}
B	-4.5	-9.5	0.55
I	-5.9	-7.9	0.55
J	-6.1	-8.2	0.55
H	-6.3	-8.4	0.55
K	-6.4	-8.4	0.55

Table B.2: Input Tully-Fisher parameters for the first loop of determination of the plateau: slope, zero-point, and scatter. These parameters have been determined in a previous KLUN study, made on a larger sample (Hanski et al. 2004, [7]).

Bibliography

[1] P. Teerikorpi, *Observational selection bias affecting the determination of the extragalactic distance scale*, Annual Review of Astronomy and Astrophysics 1997, 35:101-36.

[2] M. Hanski, *Cosmological distances, dynamics, and distributions : A Journey to* H_0 *and beyond via 6619 Spiral galaxies*, PhD thesis, Turku University (2001).

[3] `http://salish.dao.nrc.ca:8080/astrocat/courteau_help.html`

[4] G. Theureau, M. Hanski, P. Teerikorpi, L. Bottinelli, T. Ekholm, L. Gouguenheim and G. Paturel, *Kinematics of the local universe : IV. Type dependance in the diameter Tully-Fisher relation and implications on the mass-luminosity structure*, Astronomy and Astrophysics 319, 435-449 (1997).

[5] G. Theureau, M. Hanski, T. Ekholm, L. Bottinelli, L. Gouguenheim, G. Paturel, and P. Teerikorpi, *Kinematics of the local universe : V. The value of* H_0 *from the Tully-Fisher* B *and* $\log D_{25}$ *relations for field galaxies*, Astronomy and Astrophysics 322, 730-746 (1997).

[6] G. Theureau, L. Botinnelli, N. Coudreau-Durand, L. Gouguenheim, N. Hallet, M. Loulergue, G. Paturel, and P. Teerikorpi, *Kinematics of the local universe : VII. New 21-cm line measurements of 2112 galaxies*, Astronomy and Astrophysics Suppl. Ser. 130, 333-339 (1998).

[7] M. Hanski, G. Theureau, and G. Paturel, *KLUN+ Peculiar velocity catalog based on Tully-Fisher and Faber-Jackson distances*, Astronomy and Astrophysics, in press (2004).

[8] `http://cas.sdss.org/dr2/en/proj/basic/galaxies/tuningfork.asp`

[9] G. Paturel, G. Theureau, L. Bottinelli, L. Gouguenheim, N. Coudreau-Durand, N. Hallet and C. Petit, *HYPERLEDA : II. The homogenized H I data*, Astronomy and Astrophysics 412, 57-67 (2003).

[10] G. Paturel, G. Theureau, P. Fouqué, J.N. Terry, I. Musella and T. Ekholm, *Calibration of the distance scale from galactic Cepheids: I. Calibration based on the GFG sample*, Astronomy and Astrophysics 383, 398-409 (2002).

[11] L. Bottinelli, L. Gouguenheim, G. Paturel and P. Teerikorpi, *The Malmquist bias and the value of* H_0 *from the Tully-Fisher relation*, Astronomy and Astrophysics 156, 157-171 (1986).

[12] J.A. Willick, “Measurement of galaxy distances”, *Formation of Structure in the Universe*, Eds A. Dekel and J. Ostriker, 1999, Cambridge University Press.

[13] Sidney van der Bergh, *Galaxy Morphology and Classification*, 1999, Cambridge University Press.