The Extragalactic Distance Scale and the Hubble Constant

L. Bottinelli and L. Gouguenheim, Paris

(Observatoire de Paris and Université Paris XI)

A precise knowledge of the extragalactic distance scale is required in order to determine the physical properties of galaxies, or to map in detail their distribution in space. But the most important problem may be that related to the determination of the expansion rate of the Universe expressed by

the Hubble relation: $V_0 = H_0 D$ (1)

which holds for galaxies at distances small enough for the Euclidian model of the Universe to be applicable. The Hubble constant H_0 is the ratio of the cosmological velocity V_0 of a galaxy, measured with respect to the centre of mass of the Local Group, and the distance D from it.

In order to use relation (1) for determining H_0, one needs a precise knowledge of both distance and velocity. Radial velocities can be measured now with the rather high accuracy of a few km/s, but because of the random motion of galaxies, they provide a good measure of cosmological velocities only for galaxies far enough away for this component to be neglected. Random motion also aggravates the difficulties encountered in determining distances which, in practice, for relatively distant galaxies must be obtained through a succession of calibrations whose accuracy decreases with increasing distance, and where each step, depends on the previous one.

A "primary calibration" is defined as one obtained from a relation between two parameters of an object, belonging to our galaxy only one of which involves distance in its interpretation. An example is the period P and the mean absolute magnitude $\overline{M}$ for a cepheid star which can be deduced from its light curve. The distance D is obtained from the mean apparent magnitude $\overline{m}$ of the star (see panel 1) using the relation

$$\overline{m} = \overline{M} + 5 \log D + 25 \qquad (2)$$

(where D is expressed in Mpc). Although such primary calibrators give the distance of nearby galaxies, they fail when the distance increases because individual stars are not bright enough to be observed.

A "secondary calibration" is a relation between two physical properties of a galaxy obtained from galaxies whose distances have been derived from primary calibrators: e.g., the magnitude of the most luminous red stars in a galaxy of given morphological type. These stars have very high intrinsic luminosities and can thus be observed at large distances. However, they cannot be primary calibrators owing to their scarcity; they are missing from the very limited region around our solar system that can be observed.

From galaxies with distances determined from primary and secondary calibrators, a new relation — a "tertiary calibration" — between global properties can be obtained. This relation could be, for example, the absolute magnitude of the whole galaxy having a given morphological type.

It must be emphasized that any systematic error at a given step is carried over into the next. For example, if the distance of the nearby stellar cluster of Hyades, from which the $(P, \overline{M})$ relation of cepheid stars is calibrated, is in error by 10%, the same error applies to the whole distance scale, and also to the Hubble constant. Moreover, various biases can influence the statistical samples studied, and must be carefully investigated.

Historical Considerations

The extragalactic nature of the Andromeda galaxy, M31, was first established by E. Hubble in 1926 through the $(P, \overline{M})$ relation of cepheid stars discovered by H. Leavitt in 1912 in the Small Magellanic Cloud and then calibrated by E. Hertzsprung and H. Shapley from a small number of nearby cepheids (see panel 2). Using the most luminous blue stars in spiral galaxies as a secondary calibrator and the absolute magnitude of the whole galaxy as a tertiary, Hubble obtained in 1936 the first determination of the Hubble constant H_0 = 526 km/s Mpc, with an accuracy of 15%, as determined from the internal dispersion of individual values but not taking into account possible systematic errors.

Later, several systematic errors were discovered. First a difference in the zero-point of the $(P, \overline{M})$ relation for the two different classes of cepheids (disk stellar population I and halo stellar population II) was postulated by H. Mineur in 1944 and confirmed by W. Baade in 1952. Second, with the increasing sensitivity of photometry, A. Sandage in 1956 using the 200 in telescope of Mount Palomar showed that what Hubble had taken for individual luminous stars were in fact luminous star clusters.

Allowing for these errors a corrected value of the Hubble constant: H_0 = 100 km/s Mpc was derived in 1961 with its corresponding "Hubble time scale" H_0^{-1} equal to 10^{10} a.

In the meantime, through refinements in the theory of stellar evolution, the age of the oldest star clusters had been assessed at 2×10^{10} a, whilst cosmochronology theory based on isotopic abundances, as developed by W. Fowler, had produced a value of the age of our Galaxy of $1.5 - 2 \times 10^{10}$ a. Because in the simplest cosmological (Freeman) models which assume that the cosmological constant is equal to zero, the age of the Universe is necessarily less than H_0^{-1}, it was strongly suspected that the Hubble constant was being overestimated.

Between 1974 and 1976, Sandage and G. Tamman reconsidered all the calibrations and decided to adopt at each step the "best" calibrator. These, on their reckoning were:

- the $(P, \overline{M})$ relation of cepheids, corrected for a colour effect, as the primary calibration,
- the linear diameter of the largest HII region in a galaxy as the secondary,
- the absolute magnitude of the galaxy as a whole, determined from the "luminosity class" as assigned by S. van den Bergh based on its morphological features.

Panel 1. The Magnitude Scales

The apparent magnitude m of a celestial object is derived from the power per m² received, E (generally called "apparent brightness") using the relation:

$$m = -2.5 \log_{10} E + C$$

where C is an arbitrarily chosen constant.

The absolute magnitude M is the magnitude which would be observed at a distance of 10 parsec (30×10^{16} m) from the source.

With distance from the source, d, measured in pc $E = E_{10}(d/10)^{-2}$ and $(m - M)$ known as the "distance modulus" $= 5 \log_{10} d - 5$

or $5 \log_{10} d + 25$ if d is measured in M pc.

From these calibrations and a sample of distant galaxies belonging to the luminosity I class of spiral galaxies (intrinsically the most luminous and so observable at the largest distances) they obtained what they called their "preferred value" of $H_0 = 50$ km/s Mpc which corresponds to a time scale $H_0^{-1} = 2 \times 10^{10}$ a, in agreement with the determinations of the age of the oldest objects, previously mentioned.

However, this result was criticized by several authors. L. Bottinelli and L. Gouguenheim in 1976 first brought attention to the unjustified extrapolation to high luminosity classes of the diameter relation of the largest HII region. This had been determined only for low luminosity classes of galaxies by Sandage and Tamman for use as their secondary calibrator. Because luminosity class is not a continuous physical parameter, extrapolating such a relation has no meaning. The extrapolation of the $(P, \overline{M})$ relation to very luminous cepheid stars and the difficult problem of galactic extinction of light was then tackled by G. de Vaucouleurs who published in 1978 and 1979 a new discussion of the various calibrations and the distance scale, leading again to the value $H_0 = 100$ km/s Mpc. The main strength of his method lay in the use made of the largest possible variety of calibrators (five or six at each step), which should minimize the effects of hidden biases.

Radioastronomical Method

The radioastronomical method relies on an empirical determination of the relation between mass m and luminosity L of a spiral galaxy. The rotation curve of a disk galaxy, mapping the circular velocity in the disk as a function of distance from the centre, reaches its maximum value V_m at a distance r_m which is directly related to the total mass of the galaxy. This implies a relation between V_m and the absolute magnitude of the galaxy.

Unfortunately, the connecting coefficient is not known *a priori*, because the masses of individual galaxies have not been determined with high accuracy although the shape of the relation can be guessed from simple models. The point mass model gives $m = r_m V_m^2/G$, where G is the universal constant of gravitation, whereas the more complex models developed by J.C. Brandt and M.J. Belton, which take account of mass distribution in a galaxy, lead to

$$m = (3/2)^{3/n}\, r_m\, V_m^2 \,/\, G\,;$$

n, which defines a family of curves, might depend slightly on the morphological type (T) as might also r_m which corresponds to the maximum on the rotation curve and is of the order of 0.1 times the photometric diameter.

Thus, one can expect a relation of the type:

$$m = f_1\,(T)\,V_m^2$$

where f_1 is a structural parameter, most probably slightly related to the galaxy type.

On the other hand a mass-luminosity relation $m/L = f_2\,(T)$ is also expected, leading to:

$$-M = a \log V_m + b \qquad (3)$$

where a and b might be functions of the morphological type. This last relation is a distance indicator because V_m can be measured without making assumptions about distance. It gives the absolute magnitude M of the galaxy and thus the distance modulus from the apparent magnitude m, through relation (2). The method, first developed by B. Tully and R. Fischer, and known as the "Tully-Fischer" or "TF" relation is powerful because the 21-cm line of neutral hydrogen, observed with a radiotelescope gives a way of determining V_m. Whereas the whole galaxy is observed as a point source by the telescope, the hydrogen spread through the galaxy gives rise to an emission line broadened by the Doppler-Fizeau effect; the line width, corrected for inclination of the disk relative to the line of sight, is roughly equal to $2\,V_m$.

The TF method has recently been refined by several authors. M. Aaronson, J. Huchra and J. Mould have used infrared apparent magnitudes, thus avoiding the difficulties in correcting properly for light extinction; Bottinelli, Gouguenheim, G. Paturel and de Vaucouleurs have corrected the observed line widths for random motions, superimposed within a galaxy on the rotational motions. Most of these recent studies have been based on sets of data comprising several hundred observations made with high sensitivity radiotelescopes (such as Arecibo, Green Bank or Nançay). They have led to an extragalactic distance scale in good agreement with de Vaucouleurs'. When proper allowance is made for the Local Supercluster (panel 3), they lead to the value $H_0 = 100$ km/s Mpc using the zero point scale (primary and secondary calibration) of de Vaucouleurs, whereas the value $H_0 = 80$ km/s Mpc is obtained when using that of Sandage and Tamman.

Present Studies on Possible Hidden Bias

In their most recent work, Bottinelli, Gouguenheim, Paturel, P. Teerikorpi and de Vaucouleurs have focussed on a detailed study of possible hidden bias. The first concerns the TF relation (3) and the possible type - (or other parameter) dependence of the two coefficients (slope and zero-point). In the sample of primary and secondary calibrators from which relation (3) was determined, no significant type - dependence was brought to light, but owing to the limited number of calibrators available, such an effect could still be present. To check on this the TF relation is being tested on specific

Panel 2. Cepheid Variable Stars

Cepheid variables are large yellow stars named after the prototype of the group δ Cephei. The variability of Cephei was discovered in 1784 by J. Goodricke; the apparent magnitude varies between 3.6 and 4.3 in a period of 5.4 days. Several hundred cepheid variables are known in our Galaxy, most of which have periods in the range 3 to 50 days and median absolute magnitudes from −1.5 to −5.

The relation between their period of light variation and their absolute magnitude at median light was discovered in 1912 by H. Leavitt. From 25 cepheids observed in the Small Magellanic Cloud (which is known to be a neighbouring galaxy and was taken in those years to be a star cluster), she discovered that the periods of the stars were related to their relative brightness (or apparent magnitude), the brighter stars having the longer periods. Since the 25 stars are all in the same stellar system, they are all at about the same distance and their relative apparent brightnesses, therefore, indicate their relative actual luminosities. To calibrate the relationship between them, the absolute magnitude of at least one cepheid variable in our vicinity was needed. This was measured for the first time by H. Shapley, in the 1930's.

Unfortunately, the next step, the determination of the distance to a cepheid variable is difficult. First, there is no cepheid near enough for its trigonometric parallax to be measured: statistical (and thus less accurate) methods must be employed, involving the proper motions and the radial velocities of these stars. Second, most of the cepheids in our Galaxy lie close to the galactic plane, where clouds of interstellar dust heavily obscure their light: the measured apparent magnitudes of these cepheids must be corrected for this extinction and these corrections are uncertain. A promising method which is not dependent on these corrections is currently being developed by B.F. Madore, using IR magnitudes; the IR light is not affected by the interstellar dust.

samples of galaxies which are homologues of primary calibrators. Such galaxies could be expected to have similar luminosities and absolute magnitudes.

A second possible source of hidden bias has to do with the limitation on apparent magnitude. Most of the 21-cm line surveys are based on optical catalogues which have a cut-off at a certain apparent magnitude so that only the intrinsically most luminous galaxies are selected at large distances. By attributing to these galaxies the mean values of the calibrators, one underestimates their luminosities and, consequently, their distances. The general tendency therefore is for the larger distances to suffer from the largest underestimations.

When using a kinematical model of the Local Supercluster for determining an unbiased value D' of the distance and the homologue sample, it appears that the Hubble constant H_0 determined from the TF relation increases with D', as expected, for D' larger than a threshold value. At smaller distances, the luminosity function of galaxies is well sampled by the data and if measurements are restricted to these nearby galaxies, the value of H_0 is found to be equal to 70 km/s Mpc, with a relatively small scatter. It must be emphasized that the kinematical model of the Local Supercluster is used only to compensate the bias by estimating the restriction in the sample.

Panel 3. Local Supercluster of Galaxies

The concept of a hierarchical structure of the Universe containing successive systems with increasing dimension and population was first worked out by Charlier (1908, 1922) in his attempt to solve the "Olbers paradox", *i.e.* the dark night problem. On studying the number density distribution of "nebulae" — still not recognized as "extragalactic nebulae" or galaxies — he noted a large excess in the north galactic hemisphere and explained this by a second order clustering. This was the first indication of the Local Supercluster of galaxies (LS). Further quantitative analyses of galaxy counts (Homlberg, 1937; Reiz, 1941) clearly established the reality of a local density excess, but the concept of a statistically isotropic and homogeneous Universe was so strong that the LS became a generally accepted reality only forty years later. Further evidence for superclustering on a scale of several 10 Mpc has also been obtained, notably from the analysis of rich clusters (Abell, 1958).

The striking concentration of bright galaxies along a great circle through the sky has been explained in detail by de Vaucouleurs (1953, 1956) as a "Milky Way" of galaxies consisting of a large flattened supersystem — or "local supergalaxy" — which is centred near the main concentration of the Virgo cluster and includes our own Galaxy and Local Group in an outlying position. The equator of this supergalaxy is nearly perpendicular (84°) to that of our own galaxy; its diameter is about 30 - 40 Mpc (if the Hubble constant H_o = 100 km/s Mpc) and our distance to the Virgo cluster is about 16 Mpc. The mass concentration in the LS is expected to induce some departures from pure Hubble flow of uniform isotropic expansion and give rise to an appreciable virgocentric velocity and a slowing down of the general expansion. From a detailed analysis of radial velocities, de Vaucouleurs (1958) has demonstrated that the local velocity field is actually non-linear and anisotropic which is consistent with a model of an LS undergoing differential expansion and rotation.

Recent studies based on several hundred galaxies have reinforced the evidence for local anisotropy of the Hubble flow (David and Peebles, 1983; de Vaucouleurs, 1984) with an infall velocity of the Local Group towards Virgo in the range 200-400 km/s. Interpretation in terms of the structure and dynamics of the LS is not yet satisfactory: although spherical (non-rotating and rotating) models (Silk, 1974; Peebles, 1976) with an r^{-n} density profile and an induced infall velocity which varies as r^{1-n} are a useful first approximation they are too simple to account in detail for the observed velocity field.

Independent Calibrations of the Distance Scale

It must be emphasized that the value of H_0 still relies on the relatively small number of primary and secondary calibrators and the distances adopted for these nearby galaxies are still subject to revision. For example, Sandage and Tamman have recently reassessed the distance moduli of M81 and M33 which have been modified respectively by 0.97 and 0.47 magnitudes. It is thus important to have independent calibrations.

Two methods have recently been proposed. One is based on the assumption that the radiation from supernovae $B_\nu(T)$ is described by the blackbody law. The principle of the method is as follows: at time t the flux f_ν (i.e. power/m^2 received at frequency ν) is observed; this is related to the distance D, the radius of the photosphere R_{ph} and the expansion velocity of the photosphere v_{ph} by the equation:

$$4\pi D^2 f_\nu = 4\pi R_{ph}^2 \, \pi B\nu(T)$$

where $R_{ph} = R_0 + v_{ph}(t-t_0)$

The values of R_0, f_ν, T and v_{ph} are determined from the spectrum of the supernova observed at time t and the time t_0 of the explosion.

The method, developed by several authors, including D. Arnett, D. Branch, R. Kirchner and R. Wagoner at first gave low values of H_0, of the order of 40 or 50 km/s Mpc. More recently, Arnett has obtained the larger value of 70 km/s Mpc and Wagoner has shown that the problem of diffusion in the atmosphere of the supernova leads to an underestimation of H_0. The most critical problem however concerns the blackbody approximation which does not really hold. Consequently, the results obtained from this method must be considered with caution.

In a second approach, de Vaucouleurs has recently used the global parameters of our own Galaxy to calibrate directly several tertiary relations, such as the TF. He finds good agreement with the primary and secondary calibrators that he had previously used.

Conclusion

The large number of studies of these past years has shown the difficulties encountered in determining the extragalactic distance scale. First, the absolute calibration of the scale depends on a relatively small number of nearby galaxies; second, we have the problem of the non-linearity of the distance scales at large distances. This non-linearity is not always brought to light when comparing scales deduced from different distance criteria, because the luminosity bias has rougly the same effect: at large distances, only the bright end of the luminosity function of galaxies is observed.

With the data presently available, it seems that the unbiased value of H_0, in the vicinity of our Local Group of galaxies (but outside the Local Supercluster of galaxies) is equal to 70 km/s Mpc.

To improve this result and to get a much more global value of H_0, we need first more accurate information on the distances of nearby galaxies and second, a considerably larger set of data, leading to better sampling of the luminosity function of galaxies at large distances.

If this value of H_0 and an age of 15-10 Ga for the oldest objects in the Universe are confirmed, the simplest cosmological models of Friedmann will no longer stand up: the cosmological constant then has to be different from zero.