15

In Hubble's shadow: early research on the expansion of the universe

Hilmar W. Duerbeck
University of Brussels (VUB)
Pleinlaan 2
1050 Brussels
Belgium

Waltraut C. Seitter
Muenster University
48149 Muenster
Germany

Abstract

An overview on the progress of theoretical and observational cosmology in the first half of the 20th century is given. We outline the Einstein, de Sitter and Friedmann-Lemaître models, and describe the quest for the observational confirmation of the de Sitter universe, as well as the first theoretical and observational work on the Friedmann-Lemaître universe. We analyze the attempts to determine the expansion parameter, and trace early research on the deceleration parameter.

15.1 Relativistic cosmology and the redshift

15.1.1 The beginnings

In 1917, about a year after the introduction of *General Relativity* (GR), Einstein [1] and de Sitter [2] published papers on the application of GR to the universe. Encountering problems in finding meaningful boundary conditions at infinity, Einstein had the remarkable idea to circumvent the problem by envisioning space as curved and closed. Both Einstein and de Sitter assumed that the universe is constant in size, a hypothesis which was acceptable at that time. Einstein assumed that only space is curved; de Sitter took time as completely equivalent to space, and assumed a constant curvature of space-time. Depending on the choice of coordinates, de Sitter's model permits different interpretations. His original solution was written in time-independent, static coordinates, which leads to the following phenomena: (1) time flows with different speed at different points in space, and (2) there exists an 'event' horizon, where the time flow becomes infinitely slow.

Both the Einstein and the de Sitter universe included a universal constant, which had not been necessary in original GR, but which is, however, needed in 'static' relativistic universes. This cosmological constant Λ corresponds to an energy density, which is not connected with matter, and thus must be a property of the vacuum. The nature of Λ as *vacuum energy density* was already recognized by de Sitter in 1917, and put on more solid ground by Lemaître [3] in 1934. Today we assume that vacuum is a quantum vacuum, whose energy density plays an important role in the formation and early history of the universe (and possibly up to the present time).

In 1922, Friedmann [4] showed that there are no further solutions of the equations of GR including Λ, which describe universes which are constant in time, but he suggested other interesting solutions which will be discussed later.

15.1.2 No cosmological redshift in the Einstein universe

Einstein's universe is based on the assumption that matter is distributed homogeneously on large scales, and does not show large-scale motions. This was the first formulation of the cosmological principle of general homogeneity and isotropy. No large-scale variations of gravitation exist, and no redshift due to gravity does occur. The total sum of masses is – in Mach's sense – the cause of inertia. The assumption of small velocities takes care of the fact that also the second relativistic spectral line shift, the Doppler effect, which had been formulated precisely in *Special Relativity*, remains small: Einstein's universe is a world without cosmological redshifts.

The spatial geometry of Einstein's world model can be visualized as a positively curved hypersphere, a three-dimensional analog of a spherical surface, of limited extent, but without boundary. The curvature of space and the curvature radius R as a measure of its extent are both constant. The curvature of Einstein's matter-filled universe is determined by the invariant matter-density, or by the positive cosmological constant that is necessary to keep it from collapsing. The larger the matter density, the smaller is the curvature radius. The cosmological constant counteracts the gravitational attraction of the masses; it avoids a gravitational collapse of the universe.

Time as the fourth dimension flows with equal speed at all points. In five-dimensional euclidean space, the space-time-continuum appears as a four-dimensional cylinder which extends in the time-axis from minus infinity to plus infinity (the Einstein universe is synonymous with the cylindrical universe). Each light ray describes a spiral curve on the surface of the four-dimensional cylinder. Depending on distance (and thus light-time) we observe today sections of time-invariable three-dimensional space (the hypersphere), how they appear at different times. If we would be able to determine distances of objects associated with these sections, and if these would not undergo temporal changes, we would be able to have a look at the complete history of the universe, as a sum of all sections. But such a task would be extraordinarily difficult – in Einstein's universe, there is no redshift that can be used as a distance indicator.

15.1.3 Redshift in the de Sitter universe

The geometry of the de Sitter universe is conceptually more complicated than the Einstein universe, since not only the three space coordinates, but also the time coordinate is included into the curvature of space-time: this space-time has a higher symmetry than the space-time in Einstein's universe. In five-dimensional euclidean space, the space-time of constant curvature appears as a four-dimensional one-shell hyperboloid, which extends for each observer from minus infinity to plus infinity.

de Sitter postulated a matter-free universe, i.e. matter-density is equal to zero; only the cosmological constant (energy-density) is present and causes the curvature. He assumed the galaxies as 'probe particles' whose masses are too small to influence the curvature of space-time. While in the matter-filled Einstein universe, collapse can only be stopped by introducing the cosmological constant, the de Sitter universe, with its positive cosmological constant and negligible amount of matter, possesses an overwhelming tendency to expand, forcing the galaxies to move away from each other and from the observer. The galaxy 'particles' in the de Sitter universe appear to be in accelerated motion away from the observer. de Sitter re-introduced the Doppler effect into cosmology.

Two types of redshift exist in the de Sitter universe: one is the Doppler effect, which occurs between the dispersing galaxies, another one is the cosmological redshift which is based on the structure of space-time. If the real world would be a de Sitter universe, the combination of both effects should be observable in the spectra of distant celestial objects. In the following sections, the *de Sitter effect* will encompass the sum of both redshifts. The de Sitter effect should increase in a *non-linear* way with distance from the observer.

In 1917, the extragalactic nature of spiral nebulae was not yet proven. Nevertheless, these objects were de Sitter's best candidates for large redshifts. His assumption was supported by Slipher's radial velocity measurements of three spiral nebulae, which had just come to his attention. de Sitter [2] wrote:

> If [...] continued observation should confirm the fact that the spiral nebulae have systematically positive radial velocities, this would certainly be an indication to adopt the hypothesis B (de Sitter-Universe) in preference to A (Einstein-Universe).

15.1.4 Alternative interpretations of redshift in the de Sitter universe

In 1918 Einstein [5] criticized the de Sitter model in pointing out that the de Sitter universe is not empty, and that it has a horizon. This horizon is analoguous to the horizon in the vicinity of a point mass, which was studied, in 1916, by Karl Schwarzschild [6] in the framework of GR. Einstein wrote:

> Das de Sittersche System dürfte also keineswegs dem Falle einer materielosen Welt, sondern vielmehr dem Falle einer Welt entsprechen, deren Materie ganz in der [Horizont-] Fläche ... konzentriert ist.[1]

Einstein's interpretation introduced an *outer* matter horizon, defined in analogy to the *inner* Schwarzschild horizon of a spherical mass distribution. For a remote observer, time flows slowly in the vicinity of the outer mass horizon of a black hole, as well as near the inner mass horizon of the universe. Einstein interpreted the slowing down of clocks (redshift of wavelengths) towards the horizon in the de Sitter universe as a gravitational redshift: The

[1] de Sitter's system should by no means correspond to a mass-less world, but to a world whose matter is concentrated in the [horizon] surface.

'mass horizon' was now – in agreement with Mach's original view – responsible for inertia. This concept, however, led to a paradox. Two observers in neighboring galaxies would see 'their' mass horizon at different places. Such a description of the world, in which the observer is at rest, the horizon is at a certain distance, and time is running more and more slowly towards the horizon, is only a *local* one. A *global* description of the de Sitter universe must look different.

Another mathematically, physically, and cosmologically more satisfying interpretation of the de Sitter universe was found by discarding the concept of a static three-dimensional space, and by assuming that space is expanding. This solution was found by Lanczos [7] in 1922, a few months after Friedmann [4] had published a general solution for stationary (spatially invariable) and non-stationary (spatially variable) universes. One year later, Lanczos [8] gave a detailed description of his de Sitter model.

Lanczos, inspired by a paper by Weyl, found that the geometrical properties of the de Sitter universe remain constant when the (positive) curvature of space is assumed to change with time. Time itself is, like in the Einstein universe, perpendicular to space. In contrast to the static de Sitter coordinates, in which time for each observer flows as a function of distance (and fastest for his own world line), Lanczos introduced a coordinate system which is defined in the whole universe (also beyond the horizon of the observer), which contracts towards the past, and expands towards the future, in which the particles are at rest with respect to the coordinates, and in which time flows at the same pace everywhere (so that a cosmic time t can be defined). The constant curvature radius R of the Einstein and de Sitter universes was replaced by Lanczos by the variable scale factor $R(t)$

$$R(t) = \frac{\mathrm{e}^t + \mathrm{e}^{-t}}{2} = \cosh t \tag{1}$$

which describes the temporal change of the coordinate system.

The distance between two points χR is given by the momentaneous scale factor $R(t)$, which at time $t = 0$ had a minimum value of 1, and the temporarily independent coordinate distance χ between any two points. Lanczos [9] wrote

> Diese Kontraktion und Expansion des Krümmungsradius mit der Zeit würde übrigens unserer unmittelbaren Beobachtung insofern entzogen sein, als auch unsere Längenmaße im selben Sinne verändert werden... Die Welt würde einen 'statischen' Eindruck machen, solange keine Geschehnisse in Frage kommen – z.B. Lichtsignale – die einen Vergleich zwischen zwei verschiedenen Zeitpunkten ermöglichen.[2]

[2]This temporal contraction and expansion of the curvature radius would not be directly

Of course, these light signals are from distant galaxies. They were emitted when the universe had a different scale factor. With their help we can read the change of distances on cosmic scales (thus in Lanczos' view laboratory scales participate in the cosmic expansion).

Two years later, in 1925, Lemaître [9] made a new attempt to separate time and space in the de Sitter universe by means of another coordinate transformation. He also introduced a 'cosmic' time t and found an expanding universe, where the world line of the observer is not a preferred one any more, and where the scale factor in the past (or in the future) goes through zero. The concept of an expanding world was not unplausible for Lemaître, but he was worried that space was not curved any more after the transformation, but was flat, so that all the problems of the boundary conditions of the world at infinity, which had been removed by Einstein and de Sitter by assuming a closed universe or a universe with a horizon, resurfaced again.

Finally, in 1928, Robertson [10] suggested a similar coordinate transformation for the de Sitter universe, and derived a redshift-distance relation that was, at first approximation, linear.

15.1.5 Wavelength shifts in the models of Friedmann (1922) and Lemaître (1927)

Both Friedmann [4], in 1922, and Lemaître [11], in 1927, first described nonstationary universe of positive curvature. Besides the mass density and the cosmological constant, Lemaître also took into account the pressure of matter and of the vacuum. In 1924, Friedmann suggested the possibility of a negatively curved universe [12]. The flat (euclidean) *Friedmann-Lemaître universe* (FL-universe) was discussed in in 1929 by Robertson [13] and in 1931 by Heckmann [14]. Curvature in FL-universes is described by the time-dependent scale factor $R(t)$ and the curvature factor k, which is constant and has the values $-1, 0, +1$ in open, flat and closed universes.

If local gravitational perturbations caused by the peculiar motions of masses are neglected, all masses in the FL-universe (as well as in the Lanczos-universe) keep the same coordinate distances for all times, while space, which carries the coordinate system, does change. In an expanding universe, a redshift occurs which increases with distance of the source from the observer. A blueshift would be observable if the universe contracts. Only observation can tell in which direction the spectral lines are shifted. Theory does not make any prediction about the sign of displacement.

---

observable, since our length scales change in the same way ... The world would appear static, unless no events – e.g. light signals – are considered, which permit a comparison between two different moments.

The Einstein and the de Sitter universes are limiting cases of the FL-universe. In general, three-dimensional space changes while time is perpendicular to it.[3] Expanding space has an initial value, which may be zero or positive, and may expand to infinity. Space can also periodically expand and contract. Whether it expands forever or contracts at some time depends on the preponderance of the repulsive force of a positive cosmological constant or the attraction of the masses.

In FL-universes, the hypersphere is the locus of equal (cosmic) time. All informations reaching us can be referred to the actual (or another) hypersphere, if the distances of the light sources are known. Distances in FL-universes can be calculated from the parameters that determine k and the function $R(t)$, if we assume that the light sources do not suffer any unknown, time-dependent changes. In such universes, relatively small Doppler shifts (from peculiar motions) and gravitational redshifts (which become non-negligible only in the vicinity of black holes) are superimposed on the cosmological redshift, which dominates at large distances. Since about 1930, this space is the topic of research of observational cosmology.

One assumption has been made for all described universes: the principle of the constancy of the velocity of light, which is also the foundation of *Special Relativity*. It permits to measure the expansion of space.

15.2 Searching for the de Sitter effect

15.2.1 Research in the years 1920–1930

Irrespective of the character of the de Sitter universe, whether it is a stationary or expanding world, many astronomers in the 1920s searched for motions of the spiral nebulae, and found in their results a confirmation of the de Sitter model.

The line of observers starts in 1924 with Wirtz, Silberstein and Lundmark, and culminates in 1929 with Hubble, who announced a linear relation between distance and redshift of galaxies. This is the famous Hubble law, which he introduced with the following words:

> the outstanding feature ... is the possibility that the velocity-distance relation may represent the de Sitter effect, and hence that numerical data may be introduced into discussions of the general curvature of space

[3]This assumption is not necessary, and was introduced by Friedmann only to simplify the mathematical treatment.

and

> it may be emphasized that the linear relation found in the present discussion is a first approximation representing a restricted range in distance.

Hubble was the *last* of the early cosmologists who believed that his result confirmed the de Sitter model.

Hubble has been put into the foreground by a glorifying historiography in such a way that astronomers who had in their hands equally large parts of the truth (or better, confirmation of a model by measurements) now live a shadowy existence. In this article, it is attempted to present the observational data and the theoretical basis of the various researchers. Using the original graphical relations (sometimes reconstructed from quoted sources and shown here for the first time), we can get an idea of the insight and perspective of these researchers. At least one of them, Lemaître, has supported his own theoretical model by observational data – already in 1927: he found the observable expanding universe. If one adds the fact that Lemaître [15], in 1931, took a physically founded choice among the possible non-stationary universes by suggesting the decay of a primeval atom (the *big bang*, as it was later baptized by Fred Hoyle), he really is the true 'father of the expanding universe'.

15.2.2 The radial velocities of nebulae

In 1917, when de Sitter had only three data at hand, already 25 radial velocity measurements of spirals were known [16]. It seems, however, that they were not yet available in de Sitter's working place, the Leiden Observatory. Six years later, Eddington published a list of 41 entries in his textbook *The Mathematical Theory of Relativity* [17], about which he says:

> The most extensive measurements of radial velocities of spiral nebulae have been made by Prof. V.M. Slipher at the Lowell Observatory. He has kindly prepared for me the following table, containing many unpublished results. It is believed to be complete up to date (Feb. 1922).

Until 1930, this list of radial velocities remained the fundamental data base for investigations in observational cosmology. Afterwards, more redshifts were added, mainly by Humason at Mt. Wilson Observatory. In 1958, about 800 redshifts had been observed, mainly at Mt. Wilson, Mt. Palomar and at Lick Observatories.

Why were these early radial velocity measurements made, these most important data – besides the distance determinations – for the new science of cosmology?

Slipher's work on spectra of nebulae grew from a task that was given to him by Percival Lowell at his observatory, and that was originally motivated by Lowell's interest in the solar system and its origin. Until the first years of the 20$^{\text{th}}$ century, some astronomers took spiral nebulae for whirls of matter, from which new planetary systems form according to the Kant-Laplace hypothesis. When Hartwig, in 1885, discovered a supernova in the Andromeda nebula, he heralded it with the words

> Da ist schon die Zentralsonne im Nebel[4]

Already in 1899, Keeler at Lick Observatory had recorded the continuous spectrum of the Andromeda nebula; at the same time, Scheiner at Potsdam Observatory had found traces of absorption lines in its spectrum. Slipher equipped his spectrograph with an especially light-efficient optics, with which he could record spectra of objects of low surface brightness better than anyone else. In combination with the first spectra of the Andromeda nebula, which clearly showed lines, Slipher published the spectrum of the nebula in the Pleiades, which, as a galactic reflection nebula, also showed a continuum with absorption lines. Unlike the Lick spectrograph, the Lowell spectrograph was stable enough to permit the measurement of radial velocities. The measurement of the high radial velocities, which are found in the spirals, but not in other nebulae, made the question of their unusual character emerge.

15.2.3 The K effect

In the first decades of the 20$^{\text{th}}$ century, several astronomers determined the proper motions and radial velocities of special groups of stars relative to the Sun, in order to determine their secular parallax (and thus their mean distance) as well as the space motion of the Sun. In 1910, W.W. Campbell, while determining the solar motion relative to the radial velocity of B stars, found a systematic trend which he named the *K effect*. This trend could be interpreted in different ways: by systematic errors in the wavelengths of lines used for the radial velocity determination, or by a general expansion of the system of B stars away from the Sun. Decades later, the effect found in the B stars vanished from literature; better wavelengths and the discovery of the rotation of the Milky Way had made it superfluous. But it continued to play an important role in the radial velocities of extragalactic nebulae.

[4] Here we have the central sun in the nebula [of a forming planetary system].

Since 1915, various groups in the US and Canada tried to determine the solar motion and the K effect from the radial velocities of nebulae. The very large value of K indicated a group of objects that expanded with a large velocity away from the sun. In 1918, Harlow and Martha Shapley were the first to investigate a correlation of the high velocities with other characteristics, e.g. the apparent velocities. In 1921, Wirtz, when studying the K effect, described an unpublished diagram, which showed an almost linear relation between apparent nebular magnitudes and radial velocities. Three years later Wirtz [18] found a linear relation between the radial velocities and the logarithms of the major axes (in arc seconds) of the nebulae. He interpreted his finding as an indication of the de Sitter effect. Since both the apparent magnitude and the apparent diameter are an indicator of the distance on an object, the way for investigating the de Sitter effect was clear: in addition to the radial velocities, distances to the nebulae had to be measured, and the form of the relation had to be derived. At a closer sight, Wirtz' first attempt does not look too convincing: both the use of magnitudes and of the logarithm of the diameters indicates that the redshift grows only logarithmically with distance (for Wirtz, this was a good argument against the existence of superluminal motions in the line of sight).

15.2.4 Nebulae as galaxies

Around this time, in 1923, one of the most important events of extragalactic astronomy took place: while searching for novae (which were not accepted by all astronomers as good distance indicators), Hubble discovered cepheid variable stars in the Andromeda nebula. Their quality as distance indicators was beyond doubt: in 1912, Henrietta Leavitt had derived a period-luminosity relation from the light variations of the cepheids in the Small Magellanic Cloud, which soon afterwards was calibrated by Hertzsprung. The spiral nebulae were now recognized, by means of their cepheids, as galaxies far away from our own, with sizes that sometimes rival that of the Milky Way.

In the beginning, the number of nebulae whose distances could be determined by means of cepheids and other visible member stars, remained very small. Apparent total magnitudes had to be used for the determination of most galaxy distances, assuming that the corresponding absolute magnitudes are on the average similar. An analoguous approach was based on apparent and absolute galaxy diameters.

Two catalogues with apparent nebular magnitudes were used in the 1920s for the construction of a radial velocity-distance relation. One of them is based on visual estimates by Holetschek, published in 1907. They were put to a normal scale by Hopmann [19] who made photometric measurements of Holetschek's comparison stars using a wedge photometer. The second catalogue was made by Wirtz [20], who, between 1911 and 1916, determined surface and total magnitudes of nebulae with a visual photometer attached to the large refractor of Strasbourg Observatory.

15.3 The (hidden) Hubble constant

15.3.1 The radial velocity – distance relation

Many theoretical investigations of the 1920s by Weyl, Eddington, Silberstein, Tolman, Robertson and others dealt with the question what relation between radial velocity v and distance d exists in the de Sitter universe. A general result was that $v = d^n$, with the exponent n being between 1 and 2. From observational data, Wirtz obtained at first a relation $v \propto \log d$. Lundmark fitted his data of 1924 with a polynomial; $v \propto K_1 \times d + K_2 \times d^2$. Hubble saw a convincing linear relation in his data of 1929, and concluded that he could neglect a quadratic term. For the linear part, he determined the slope that was later named the Hubble constant in his honour.

But already before Hubble's contribution, the investigations by Lundmark, Strömberg and Lemaître yield this constant. One should also not overlook the papers by Silberstein in 1924, but his radial velocity-distance diagram is not convincing for present-day readers, since it included galactic cepheids, globular clusters, and the Magellanic Clouds.

15.3.2 Lundmark: the first 'Hubble diagram'

The first radial velocity-distance diagram of galaxies was published by Lundmark [21] in 1924. The radial velocities were taken from Slipher's studies. The distances were determined from the apparent magnitudes of Holetschek-Hopmann. Assuming that galaxies have all the same absolute magnitude, he chose the Andromeda nebula as a reference galaxy and gave all distances in units of the distance of Andromeda. Using the angular diameters as a second distance indicator, again with reference to Andromeda, a second distance estimate was made. Lundmark's radial velocity-distance diagram contains 44 objects, it was described by the author as follows:

> Plotting the radial velocities against [these] relative distances, we find that there may be a relation between the two quantities, although not a very definite one.

Lundmark's data extend to quite large distances, about 40 or 100 Mpc, with his assumed values for the distance to Andromeda. The data are quite insecure, and it is not surprising that the correlation coefficient for a linear relation is only 0.37. If one assumes for the distance to the Andromeda galaxy Lundmark's smaller value of 0.2 Mpc (which he calls 'probably the best', and which he obtained from the comparison of luminosities of novae in the Milky Way – whose absolute magnitudes were approximately known – with those in the Andromeda nebula), then one gets for the slope of a straight line through the zero point and his data points the value

$$H_0 = 90 \pm 12 \text{ km/s/Mpc}\,.$$

for the Hubble constant.

Since the Hubble constant, being the inverse of the time that can be interpreted as the 'age of the universe' in some simple model, is variable with time, its present-day value is labelled with the index 0. The quadratic term, that was also included and determined by Lundmark, is neglected in the present investigation, since we want to analyse the data in a similar way as Hubble. The value that was derived by Lundmark on the basis of improved nova magnitudes of 0.5 Mpc for the Andromeda galaxy yields the following value for the Hubble constant:

$$H_0 = 36 \text{ km/s/Mpc}\,.$$

It is not clear why Lundmark in his investigation did not correct the radial velocities of the nebulae for solar motion. Corrected values are used for the newly derived Hubble constants and in Fig. 15.1.

15.3.3 From Strömberg to Hubble

In 1925, Strömberg [22] at Mt. Wilson Observatory collected a list of radial velocities, which originated with few exceptions from Slipher's observations. He used the apparent galaxy magnitudes from Wirtz's [19] catalogue[5] for the distance determinations. Strömberg also made the assumption that the absolute magnitudes of all nebulae are equal, but he dropped the problem of calibration and calculated only relative distances $r = 10^{0.2\mathrm{m}}$. The correlation coefficient for the relation between radial velocity and relative distance is 0.23. Strömberg's conclusion

> . . . we may say that we have found no sufficient reason to believe that there exists any dependence of radial motion upon distance from the sun

[5] which, according to an unpublished study by the authors, are of fairly low quality

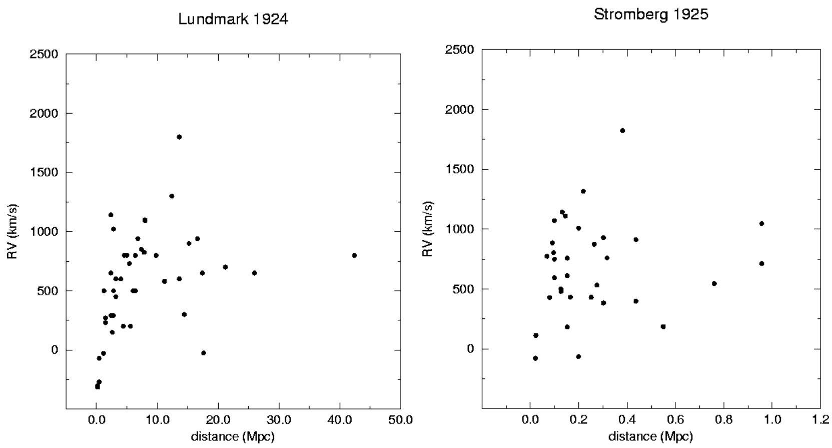

Figure 15.1: The radial velocity – distance relation of Lundmark (1924) and Strömberg (1925).

is completely understandable on the basis of his data. They yield a scatter diagram, and the Hubble constant, which has a very large error, is several times larger than those of other authors (Fig. 15.1).

Robertson [10], in his paper of 1928, derived a redshift-distance relation for the de Sitter universe for small distances:

$$v/c = r/R = \mathrm{const} \tag{3}$$

with velocity v, distance r, and velocity of light c. By inserting data (distances from Hubble [23] and radial velocities of Slipher) he derived the constant radius of curvature R. It can be seen from equation (3) that the ratio c/R corresponds to the Hubble constant. Robertson's value of R yields

$$H_0 = 460 \text{ km/s/Mpc}.$$

One year later, Hubble [24] used Slipher's radial velocities, supplemented by a few additional ones by Humason, and his newly derived distance determinations, mainly based on brightness estimates of the brightest stars in galaxies (or at least of the objects that he took for the brightest stars), to determine the Hubble constant

$$H_0 = 535 \pm 40 \text{ km/s/Mpc}.$$

The correlation coefficient of his data, 0.84, is much better than those of his predecessors. His determination was the last one in the framework of the de Sitter universe (Fig. 15.2).

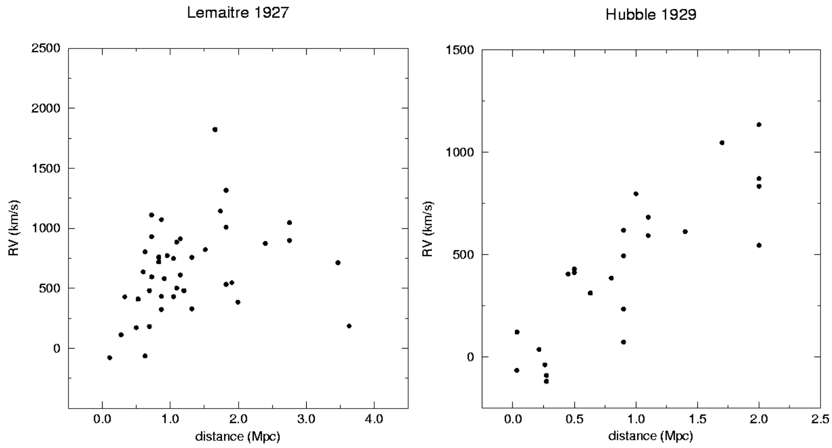

Figure 15.2: The radial velocity – distance relation of Lemaître (1927) and Hubble (1929).

15.3.4 The Hubble constant of Lemaître

In 1927, two years before Hubble's investigation, Lemaître [10] deduced the equation for the relative change of the wavelength in an expanding, matter-filled universe. Since one deals here with the expansion or contraction of the coordinate system, Lemaître used in this context the expression '*apparent* Doppler effect' and put

$$\frac{\lambda - \lambda_1}{\lambda_1} = \frac{R - R_1}{R_1} = \frac{v}{c} \tag{4}$$

with λ = observed wavelength, λ_1 = emitted wavelength, R = present-day scale factor, R_1 = scale factor at the time of the light emission. The most important difference to the *true* Doppler effect is the following: the signal-propagating motions, which lead to the true Doppler effect, cannot transgress the velocity of light. For the true Doppler effect, equation (4) is only an approximation, which has to be replaced by the Doppler formula of special relativity. In the case of the *apparent* Doppler effect, the equation is exact, and should not be replaced by the relativistic Doppler formula under any circumstances, because this would lead to completely wrong concepts about the structure of the universe.

Similar to Robertson, who derived for the case of the de Sitter universe an approximation for v/c in the case of small distances (not small velocities!)

between source and observer, Lemaître introduced a similar approximation for the expanding universe in the case of *small distances* from the observer. The difference in equation (4) can be replaced by the differential dR, and one obtains:

$$v/c = dR/R = [(dR/dt)/R]dt = (\dot{R}/R)dt = (\dot{R}/R)\rho \tag{5}$$

where ρ is the distance between observer and source in units of time, and $\dot{R}$ is the temporal derivative of the scale factor. With $\rho c = r$ (in units of length), one obtains the approximation

$$\dot{R}_0/R_0 = v/r\,. \tag{6}$$

Its present-day form is $\dot{R}_0/R_0 = H_0$, and the above-derived approximation is

$$v/r = H_0\,. \tag{7}$$

Lemaître did not stop after the theoretical deduction, but applied his results on the data which were available to him. As many others, he took Slipher's radial velocities, as they were found, supplemented by some results of others, from Strömberg's list. The galaxy distances r were derived from the apparent magnitudes, which were given in Hubble's extract of Holetschek's catalog. The mean absolute magnitude of a galaxy, $-15\overset{\mathrm{m}}{.}2$, was taken from from Hubble's paper of 1926.

In Lemaître's paper of 1927, the Hubble constant was directly determined for the first time, in today's sense of the word: as a characteristic parameter of the expanding universe. Lemaître obtained the weighted value

$$H_0 = 625 \text{ km/sec/Mpc.}$$

A correlation coefficient of 0.30 is derived from the 42 points, shown in Fig. 15.2. Lemaître's paper did not give a figure. An unweighted Hubble constant of 615 ± 70 km/s/Mpc is derived. Since we cannot retrieve Lemaître's weighting scheme, we cannot recover his deduced value.

The faint value for the average absolute magnitude of galaxies, as quoted above, is certainly the reason for the very large values of the Hubble constant deduced between 1926 and 1930, which are based more or less on the same data set. Thus there is no reason to suspect (as Peebles has done) that there were hidden relations between Lemaître and Hubble because of the similarity of their results. Most data had been published by 1925 and were available to all interested cosmologists. It is only surprising that the determination of individual distances for the 24 galaxies in Hubble's diagram of 1929 obviously has not removed the systematic error in galaxy luminosities, but has only reduced the scatter in the data points. A 1931 paper by Oort is important in this context, and will be discussed below.

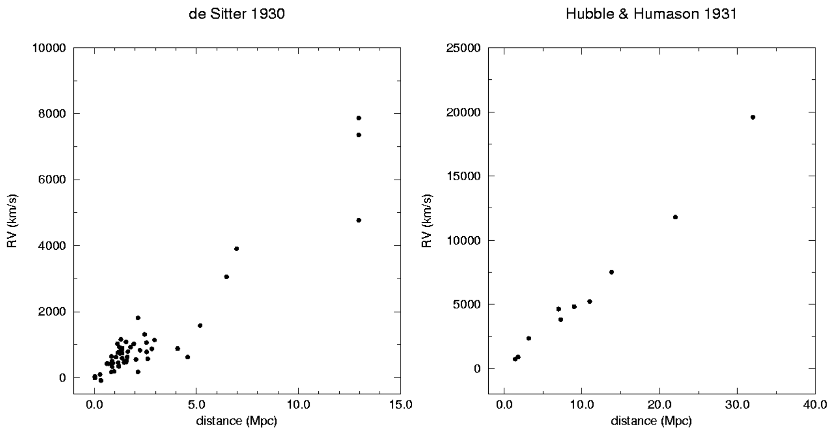

Figure 15.3: The radial velocity – distance relation of de Sitter (1930) and Hubble and Humason (1931).

15.4 The time after the 'Hubble law'

15.4.1 De Sitter and his 'Hubble relation'

In early 1930, about a year after Hubble's report, de Sitter [25] published a quantitative analysis of the redshift-distance relation. He divided the nebulae in three morphological groups: spirals, ellipticals, and irregulars. He investigated for all of them separately if the apparent nebular magnitudes and the logarithms of the apparent diameters are correlated in a linear way. The good correlations indicated that both criteria were suitable for the determination of distances. de Sitter compiled a photometric and radial-velocity catalog for 54 galaxies. Since the data extend to larger distances than before, he obtained a correlation coefficient of 0.94, which is better than that obtained by Hubble the previous year (Fig. 15.3).

The data of de Sitter yield the following Hubble constant (after transformation into Mpc, since de Sitter used the now obsolete distance unit *Andromeda* $= 10^{22}$ m):

$$H_0 = 520 \pm 20 \text{ km/s/Mpc}\,.$$

de Sitter himself quotes a value of 465 km/s/Mpc.

Hubble was obviously furious about this publication, since it preceded his first detailed paper after the short note of 1929 in the Academy proceedings. After an exchange of letters with de Sitter, a fundamental paper by Hubble and Humason [27] finally appeared in 1931, where the Hubble constant is determined to be

$$H_0 = 558 \text{ km/s/Mpc}$$

(Fig. 15.3). Hubble, in a footnote, draws attention to the numerical similarity of the value of de Sitter and the one in their paper, and explains this (with good reason) by the more or less identical data sets.

More important than de Sitter's numerical value is his insight that neither Einstein's universe (A) nor his own de Sitter universe (B) can explain the observations:

> We ... come to the conclusion that neither the solution (A) nor (B) can correspond to the truth, (A) being excluded by the large positive systematic velocity V [of galaxies], and (B) by the finite density, which is excluded by [the assumption of a massless universe].

Here, de Sitter quotes Lemaître's paper of 1927, with which he had become accustomed a few weeks before by Eddington. According to de Sitter, Lemaître's 'ingenious' expanding universe is the correct solution. This is the first public recognition of the expanding-matter universe. Max Planck once said

> Eine neue wissenschaftliche Wahrheit pflegt sich nicht in der Weise durchzusetzen, daß ihre Gegner überzeugt werden und sich als belehrt erklären, sondern vielmehr dadurch, daß die Gegner allmählich aussterben und daß die heranwachsende Generation von vorneherein mit der Wahrheit vertraut gemacht wird.[6]

What a contrast is de Sitter's reaction! Here, the inventor of a world model discards it, and praises that of someone else.

The sequence of events leading to the acceptance of the expanding universe was indeed somewhat confusing and could only with hindsight be interpreted as a triumph of Hubble. Besides de Sitter, Eddington earns prize to have propagated Lemaître's paper of 1927. When, in 1929–30, he was busy with ideas about the 'instability of Einstein's spherical world', he mentioned during a discussion following a talk by de Sitter that the problem of cosmology possibly lies in the limitation of static models [27].

Lemaître, who in 1925 had been a research student with Eddington, read the contribution to the discussion and reminded Eddington about his solution

[6] A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with that truth.

of 1927, which had obviously been overlooked completely by his former teacher (and all others). Immediately Eddington initiated an English translation of Lemaître's paper which appeared in the Monthly Notices of 1931. It remains an unsolved riddle why the determination of the Hubble constant and its value in km/s/Mpc was omitted in the translation; the only explanation is that Eddington thought that better determinations were now available, and wanted to keep Lemaître's brilliant theory separated from his poor (albeit in 1927 state-of-the-art) observational verification: 26 lines are missing, in another context another three lines, otherwise the translation is complete.

15.4.2 Oort's work of 1931

Oort is one of the pioneers of Milky Way research, but in his youth (and in his later age) he also did important extragalactic work. At the end of 1931, he published [28] a paper entitled *Some problems concerning the distribution of luminosities and peculiar velocities of extragalactic nebulae*. Here, Oort investigated the distribution of luminosities in a given volume element (the luminosity function) on the basis of the observed radial velocities and the apparent magnitudes. He found a dominance of bright galaxies in clusters and of faint galaxies in our vicinity. His researches also yielded systematic differences in the velocity dispersions. Oort pointed at the possible existence of a local galaxy cluster as the reason for the smaller velocity dispersions in our vicinity.

To compensate for the systematic differences, but especially in order to avoid the distance estimates which become more and more inaccurate with growing distance, Oort replaced in the relation

$$M = m + 5 - 5 \log r \tag{7}$$

r by v/H. Under the assumption that the average absolute magnitudes at large distances are the same as the average absolute magnitudes at small distances, he obtained

$$M = < m - 5 \log v > \ + 5 \log H - 25 \tag{8}$$

where the additive constant takes into consideration the conversion factor from pc into Mpc.

Oort described the inaccuracies of distance determinations as follows:

> Hubble and Humason advocate the use of these brightest stars for determination of the distance. They give reasons for believing that the absolute magnitude of these brightest objects is a fairly well determined quantity, about equal to $-6\overset{\mathrm{m}}{.}1$. Not having seen

> the plates it is difficult from the brief description to form a judgment of the reliability of this method. For one thing it must have been impossible in many cases to discriminate between stars and compact diffuse nebulosities or open and globular star clusters.

Indeed, Humason, Mayall and Sandage [29] proved in 1956 that the 'bright stars in distant galaxies' of Hubble are to a large extent H II regions of high absolute magnitude. This is a main reason for the too high value of the early Hubble constant.

Oort determines carefully from different morphological types the average value

$$H_0 = 290 \text{ km/sec/Mpc},$$

and suspects 'it can be in error by a factor 1.5, or possibly 2, in each direction.' Disregarding the very early estimates of Lundmark (who actually did not calculate at all something like a *Hubble constant* at that time), Oort's value of H_0 is by far the smallest for quite some time. If one takes into account that the revision of the distance scale by Cepheids in the 1950s led to an increase of a factor of two (which would mean a reduction by a factor 2 of Oort's result), his value of the Hubble constant lies astonishingly close to modern determinations (as well as the result based on Lundmark, whose distance scale, based on novae, avoided the cepheid zero point error).

15.4.3 The second temporal derivative of the scale factor

The importance of the Hubble constant, which is the first temporal derivative of the scale factor $\dot{R}$, normalized on the scale factor R, lies in its observational accessibility by the measurement of redshifts, and its importance for obtaining a first guess of the age of the universe. All other effects are much more subtle. In the mid-1930s began the search for the effect of the second temporal derivative of the scale factor $\ddot{R}$, which also occurs in the Friedmann-Lemaître equations.

In 1935 Silberstein [30] published a paper in which he polemicized that the linear law found by Hubble cannot be correct because of the gravitational interaction of galaxies (and he thus saw a reason to keep the old de Sitter model alive), and deduced – at about the same time as Tolman and de Sitter – relations between the deceleration, the matter density, and the cosmological constant.

In 1956 Hoyle and Sandage [31] defined in a joint paper the deceleration parameter q as $\ddot{R}/RH^2$. Thus the Hubble constant gained another important role: the other observational units in the Friedmann-Lemaître equations could be normalized by the square of the Hubble constant. Thus one gains from the matter density the density parameter, from Λ the Λ-parameter, and from p the pressure parameter.

Silberstein closed his 1935 paper with the polemic statement that it would take 'perhaps only after some decades of hard work' to derive numerical values for the cosmological constant and the deceleration parameter (he used equivalent parameters), 'supposed to be still inclined to the space-expansion idea'. They are, indeed, and the determination of these parameters is still a laborious task, which only now by the use of distant type Ia supernovae yields the first convincing results.

15.4.4 The multitude of homogeneous and isotropical universes

Robertson [13] deduced, from the principle of homogeneity and isotropy, the most general line element of universes, in which it obeys Einstein's field equations and permits to tie geometry with physical parameters. Robertson's results confirm once again that under the assumption of the cosmological principle, Einstein's and de Sitter's universe are the only static curved universes. Among non-stationary universes, those of positive and negative curvature are possible (as was derived by Friedmann in 1922 and 1924), and the euclidean variant of the expanding de Sitter universe, which had been found by Lemaître in 1925, and by Robertson in 1928. In addition, they include the four-dimensional euclidean Friedmann-Lemaître universe.

Explicitly the euclidean universe was first presented in Heckmann's [13] investigation of the general Friedmann-Lemaître universes (cosmological constant $\Lambda < 0, = 0, > 0$; curvature factor $k - 1, 0, +1$). Heckmann was able to show that the cosmological principle is not only a necessary, but also a sufficient condition for all Friedmann-Lemaître universes.

Heckmann's paper was published in the proceedings of the Göttingen Academy and the publications of the Göttingen Observatory. Studying Heckmann's paper, Einstein and de Sitter found the model with $\Lambda = 0$ and $k = 0$ especially interesting. Their rapidly published discussion of this model [32] lead to its name Einstein-de Sitter model. The authors justified their choice: this model is especially simple and satisfies (up to the proof of the contrary) a generally accepted principle, that of greatest simplicity. It also discarded Einstein's since long unloved cosmological constant.

The Einstein-de Sitter universe had a long lasting influence on cosmology. Till the most recent past, observational data were fitted almost exclusively by models without Λ (the so-called standard models). Since under the assumption of $\Lambda = 0$ the matter density in the universe, as derived from observations, is relatively near to the density of an euclidean universe, it appeared reasonable to assume the Einstein-de Sitter universe.

15.4.5 The inflationary universe

The Einstein-de Sitter universe gained new importance through the theory of the inflationary universe [33]. At the early stages of the universe, the cosmological constant (now in the form of the vacuum energy density) had a very large value. The high matter density could not compensate the repulsive force of Λ. Decreasing matter density made Λ dominate (as it dominates in the matter-free de Sitter universe; we therefore talk about the de Sitter phase), and the universe expanded exponentially with H in the exponent (the inflationary universe).

To stop this inflation, the 'new' cosmological constant that was constant in earlier theories had to be reduced, and the universe which had become basically empty after inflation, had to be filled with matter. In some of the recent models Λ was made to zero, and the matter density was set to the critical density which occurs in the Einstein-de Sitter universe (flat space). Since the observable mass density is several orders of magnitude lower, 'dark matter' had to be invoked. But very recent studies of distant supernovae seem to indicate that Λ has a noticeable value today, so that its contribution to the energy-mass density makes space flat, and the hypothetical dark matter – while still the main contributor of unseen mass in galaxies and clusters of galaxies – has somewhat fallen out of favour. If this route is the correct one in cosmology, only time will tell.

15.5 Epilogue of actors

What did the pioneers of theoretical and observational cosmology do, after Hubble dedicated his energy and the big instruments to extragalactic research and remained active in this field (only interruped by war research) till the end of his life?

Lundmark in 1929 became Direktor of Lund Observatory, and his scientific activity declined (but he was one of the founding fathers of ESO); de Sitter died in 1934; Wirtz in the 1930s had to struggle more and more with political difficulties and died in 1939; Strömberg and Slipher continued to work on their proper research fields, stellar statistics and planetary research.

In the 1930s, Eddington propagated the Lemaître universe; he also wrote an engaged book on the existence of the cosmological constant [34]. In later years, his interests focused more – like those of Einstein – on the search for connections between relativity and quantum mechanics. Silberstein, the ardent fighter for the de Sitter universe, published his already outdated ideas in 1930 as a book [35], and polemized against the research of Lemaître, Eddington, Hubble and Humason. The quoted paper of 1935 shows that he had insight, but not so much to recognize that he worked in a dead end street. He became one of the pioneers of optical devices for the movie and TV industry at Kodak Laboratories.

Hubble himself, in articles published in 1942 and 1953 [36, 37], became sceptic of the concept of the expanding universe, since the dimming due to redshift, as well as the large-scale distribution of nebulae, based on nebular counts, were at variance with expectations based on such a model. The observations pointed towards a universe with a strong positive curvature, and an age of the order 10^9 years which was in conflict with other age determinations of that time, and his conclusion was

> either the measures are unreliable, or red shifts do not represent expansion of the universe.

Lemaître developed in the 1930s his theory of the big bang. His concept of a primeval atom lead him to investigate the cosmic radiation (at that time mostly thought to be gamma rays) as a possible relict of the decay of such a primeval atom. We know today that his search was done under wrong premises in the wrong wavelength region. The microwave background radiation, the relic of a hot big bang, as it had been proposed in 1948 by Gamow and collaborators [38, 39], was discovered in 1965 by Penzias and Wilson [40] one year before Lemaître's death.

Bibliography

[1] Einstein, A., 1917, *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie*, SB Preuß. Akad. Wissensch. Berlin, phys.-math. Kl. No. 6.

[2] de Sitter, W., 1917, *On Einstein's theory of gravitation, and its astronomical consequences*, MNRAS 78, 3.

[3] Lemaître, G., 1934, *Evolution of the expanding universe*, Proc. Nat. Acad. Sci. Washington 20, 12.

[4] Friedmann, A., 1922, *Über die Krümmung des Raumes*, Z. Phys. 10, 377.

[5] Einstein, A., 1918, *Kritisches zu einer von Hrn. de Sitter gegebenen Lösung der Gravitationsgleichungen*, SB Preuß. Akad. Wissensch. Berlin, phys.-math. Kl., No. 12.

[6] Schwarzschild, K., 1916, *Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie*. SB Preuß. Akad. Wissensch. Berlin, phys.-math. Kl. S. 189.

[7] Lanczos, K., 1922, *Bemerkung zur de Sitterschen Welt*, Phys. Z. 23, 537.

[8] Lanczos, K., 1923, *Über die Rotverschiebung in der de Sitterschen Welt*, Z. Phys. 17, 168.

[9] Lemaître, G., 1925, *Note on de Sitter's universe*, J. Math. Phys. (MIT, Boston) 4, 188.

[10] Robertson, H.P., 1928, *On relativistic cosmology*, Phil. Mag. Ser. 7, 5, 835.

[11] Lemaître, G., 1927, *Un univers homogène de masse constante et de rayon croissant, rendant compte de la vitesse radiale des nébuleuses extragalactiques*, Ann. Soc. sci. de Bruxelles 47, Ser. A, 49.

[12] Friedmann, A., 1924, *Über die Möglichkeit einer Welt mit konstanter negativer Krümmung des Raumes*, Z. Phys. 21, 326.

[13] Robertson, H.P., 1929, *On the foundations of relativistic cosmology*, Proc. Nat. Acad. Sci. Washington 15, 822.

[14] Heckmann, O., 1932, *Die Ausdehnung der Welt in ihrer Abhängigkeit von der Zeit*, Veröff. Univ.-Sternw. Göttingen Heft 23.

[15] Lemaître, G., 1931, *The beginning of the world from the point of view of quantum theory*, Nature 127, 706.

[16] Slipher, V.M., 1917, *Nebulae*, Proc. Am. Phil. Soc. Philadelphia 56, 403.

[17] Eddington, A.S., 1923, *The mathematical theory of relativity*, London.

[18] Wirtz, C., 1924, *De Sitter's Kosmologie und die Bewegungen der Spiralnebel*, AN 222, 21.

[19] Hopmann, J., 1921, *Photometrische Untersuchungen von Nebelflecken*, AN 214, 28.

[20] Wirtz, C., 1923, *Flächenhelligkeiten von 566 Nebelflecken und Sternhaufen*, Medd. Lunds Astr. Obs. Ser. 2, No. 29.

[21] Lundmark, K., 1924, *The determination of curvature of space-time in de Sitter's world*, MNRAS 84, 747.

[22] Strömberg, G., 1925, *Analysis of radial velocities of globular clusters and non-galactic nebulae*, ApJ 61, 353.

[23] Hubble, E., 1926, *Extra-galactic nebulae*, ApJ 64, 321.

[24] Hubble, E., 1929, *A relation between distance and radial velocity among extragalactic nebulae*, Proc. Nat. Acad. Sci. 15, 168.

[25] de Sitter, W., 1930, *On the magnitudes, diameters and distances of the extragalactic nebulae, and their apparent radial velocities*, BAN 5, 157.

[26] Hubble, E., Humason, M.L., 1931, *The velocity-distance relation among extragalactic nebulae*, ApJ 74, 43.

[27] Proceedings at Meeting of the R.A.S., Obs. 53, 37f. (1930).

[28] Oort, J.H., 1931, *Some problems concerning the distribution of luminosities and peculiar velocities of extragalactic nebulae*, BAN 6, 155.

[29] Humason, M.L., Mayall, N.U., Sandage, A.R., 1956, *Redshifts and magnitudes of extragalactic nebulae*. AJ 61, 97.

[30] Silberstein, L., 1935, *On the expanding-universe doctrine*, Trans. R. Soc. Canada, Sect. 3, Ser. 3, 39, 1.

[31] Hoyle, F., Sandage, A., 19 *The second-order term in the redshift-magnitude relation*, PASP 68, 301.

[32] Einstein, A., de Sitter, W., 1932, *On the relation between the expansion and the mean density of the universe*, Proc. Nat. Acad. Sci. Washington 18, 213.

[33] Guth, A.H., 1981, *Inflationary universe: A possible solution to the horizon and flatness problems*, Phys. Rev. D 23, 347.

[34] Eddington, A., 1933, *The expanding universe*, Cambridge.

[35] Silberstein, L., 1930, *The size of the universe*, London.

[36] Hubble, E., 1942, *The problem of the expanding universe*, Amer. Scientist 30, 99.

[37] Hubble, E., 1953, *The law of red shifts*, MNRAS 113, 658.

[38] Gamow, G., 1948, *The evolution of the universe*, Nature 162, 680.

[39] Alpher, R.A., Herman, R.C., 1949, *Remarks on the evolution of the expanding universe*, Phys. Rev. 75, 1089.

[40] Penzias, A.A., Wilson, R.W., 1965, *A measurement of excess antenna temperature at 4080 Mc/s*, ApJ 142, 419