Beginnings of observational cosmology in Hubble's time: Historical overview

By ALLAN SANDAGE

Observatories of the Carnegie Institution of Washington

1. Prologue

1.1. *A History of a Name*

When the organizers of this symposium asked if I could talk on history, it was not clear if they hoped for a history of HST or a scientific history of why the telescope has been named for Hubble. There is a history to both subjects.

In the early days of planning, when the telescope was a three meter dream, it was initially called the LOT for Large Orbiting Telescope. This brought forth several objections because a cadre of adventurous astronomers had urged a site on the moon. The word "orbiting" was said to block such a plan. Consequently, the name was changed in the late 1960s to LST for Large Space Telescope.

That name was still used as late as 1974 in all the planning and in the several major symposia held in a first lobbying effort, both by industry and by the scientists, to sell the telescope. One such important symposium was held in Washington at the Sheraton-Park Hotel from January 30 to February 1, 1974. The meeting was organized by F. Peter Simmons who had become Project Manager for LST at the McDonald-Douglas Astronautics Company after his earlier role at the Grumman Aerospace Corporation as Director of Astronomy for the highly successful Orbiting Astronomical Observatories (OAO). Simmons was later to play an even larger role in coordinating and organizing a major lobbying response, both in Congressional committees and in industry in the mid 1970s, to gain support for Lyman Spitzer's (1946) early suggestion for a space telescope.

It was at the 1974 Washington symposium where much of the science and the necessary technology for the project was first publicly laid out in awesome detail. Many of the future stars of the enormously complicated project, both astronomers and engineers, spoke. The names of the industry affiliations from which the technical scientists and engineers came included Ball Brothers, Bendix, Boeing, Convair, General Dynamics, Grumman, Itek, Lockheed, Martin Marietta, McDonnell-Douglas, Perkin-Elmer, and TRW, showing the wide industry interest in the project. Representatives from NASA Headquarters and from Goddard and Marshall were also there.

Among the astronomers were Lyman Spitzer (in absentia, see Spitzer 1997), Robert O'Dell (LST project scientist), Nancy Roman (chief, astronomy/relativity, NASA), Laurence Fredrick, Margaret Burbidge, Robert Danielson, Ivan King, John Bahcall, George Herbig, Gerry Neugebauer, and Harlan Smith.

John Naugle, Associate Administrator for Space Science, NASA headquarters, gave the sobering epilogue where he outlined the major hurdles to be conquered as seen in 1974. Among the many important cautions he gave, one of the most central was: "From what you have heard over the past few days, it is quite clear that we are smart enough technically to build the Large Space Telescope now." [However] "scientists must recognize that where they are dependent upon public support for their endeavors, they must communicate the importance of their endeavors to the public—the knowledge they have gained and its importance. This enables the public to participate, in many cases vicariously, in these activities. If scientists devote perhaps one-tenth of the creative energy devoted to understanding the universe to explaining to the public the reasons for and the importance of what they are doing, then I think the problems that we have in obtaining support for basic research will disappear."

One of the grand purposes of the present workshop is to do just that.

For various reasons, mostly to do with the arcane art of political persuasion, the name of the telescope was again changed simply to ST when the aperture was reduced to 2.4 meters in the late 1970s. The rationale given was that the word "large" was too strong, suggesting not only an ultimate instrument but also an ultimate price, thereby possibly jeopardizing a future really big space telescope. However, the change of name was again opposed by those who argued that LST was the appropriate name, standing as it did for the Lyman Spitzer Telescope. The dream might not have become reality without Spitzer's vision, and of course, also not without the near ineffable genius of the engineers and scientists and the remarkable ability of industry. This symposium is for all of you who have made it possible for us.

1.2. *Hubble's legacy*

Why then was the telescope eventually named for Edwin Hubble? That too is appropriate, but much farther back in history. Simply, Hubble had manufactured the foundations upon which a large part of the present work of the telescope on cosmology is centered.

In only 12 years from 1924 to 1936, Hubble brought to an almost modern maturity the four foundations of observational cosmology, even as its principles are practiced today.

(1) He proved that nebulae are galaxies by identifying the content of NGC 6822, M33, and M31 (Hubble 1925, 1926a, 1929a) to be stars similar to those in the Milky Way.

(2) From an early beginning in 1922, he perfected the galaxy classification system (Hubble 1926b, 1936c) that undoubtedly contains clues to galaxy formation and evolution. Hubble's proposal, now universally adopted, was more systematic than that of Lundmark (1926, 1927), but there are obvious similarities, especially as to names. Lundmark introduced three groups as "amorphous ellipticals," "true spiral," and "magellanic cloud types." A flavor of a rivalry between these two giants is seen in Lundmark's (1927) footnote rebutting Hubble's (1926b) perhaps unjustified attack on Lundmark, also in a footnote.

(3) He organized existing data on redshifts and apparent magnitudes (Hubble 1929b) of nearby galaxies into a believable redshift-distance relation, searched for throughout the 1920s as the "de Sitter effect" by many others (Wertz [the European Hubble without a telescope], Truman, Silberstein, Lundmark) but without success, and seen in the early data as adumbrations by Lemaitre (1927, 1931) and Robertson (1928). Hubble, with Humason, then greatly extended the velocity-distance relation into the "remote" expansion field (Hubble & Humason 1931, 1934: Humason 1936; Hubble 1936a, 1937, 1953).

(4) He made a massive observational program of galaxy counts for the N(m) function, from which he attempted to measure the curvature of space (Hubble 1934, 1936b,c, 1937, 1953).

More detail on the history of these developments is the subject of this review. Most emphasis is placed on galaxy counts (item 4) as buttressed by data on redshifts and magnitudes (item 3) as needed for the interpretation. A few comments on the role of the abnormal galaxy morphology at faint magnitudes in the HDF, (item 2), closes the review.

2. The 1934–1936 N(m) count campaign

2.1. *The observational data from the 1934 campaign*

Building on the work of Fath (1914) as analyzed by Seares (1925), work that was based on galaxy counts using 60-inch reflector plates taken for The Mount Wilson Catalogue of Photographic Magnitudes in Selected Areas 1–139 (Seares, Kapteyn, & van Rhijn, 1930, hereafter the Mount Wilson Catalog), Hubble (1926b) used all existing data to show that the "white nebulae" increased in number as log N(m)$\sim$0.6m. This is the requirement for a uniform (homogeneous) distribution in depth in Euclidean space, regardless of any form of the distribution of absolute magnitudes (the luminosity function) as long as the integral of that function over luminosities is finite and if there are no effects on the magnitudes with distance (absorption, redshift, etc.). It is also the expected form in the limit of zero redshift, even using the modern (Mattig) equations that correctly describe the distribution (section 4).

The observational data available in 1926 was spotty and not well calibrated in magnitudes. Beginning in 1927, Hubble undertook a major survey with the Mount Wilson 60 and 100-inch telescopes to carry the survey of galaxies at increasing distances by extending the magnitude coverage beyond $m_{pg} = 16.7$ which was the effective limit of his 1926 study.

Hubble (1934) completed the massive observational program in 1934 in which he counted 44000 galaxies in an area of 650 square degrees on 1283 plates in a systematic sampling in both Galactic hemispheres. The result was a definitive study of the average properties of galaxy distribution, both in depth (for homogeneity) and around a significant fraction of the sky (for isotropy). In this major paper, Hubble (a) confirmed that galaxies continue to increase in numbers to the faintest limits surveyed (the ultimate organizational hierarchy appeared to have been reached), (b) there is a strong latitude effect showing absorption by the Galaxy, (c) the "zone of avoidance" is mapped in greater detail than was possible by Seares (1925), (d) the frequency distribution of the numbers of galaxies per square degree, when the counts on each of the 1283 plates were reduced to standard conditions (for the latitude effect, for different exposure times for the plates, for different seeing, for distance-to-center of each plate due to coma, etc), shows a normal error (Gaussian) distribution in log N, *not in N itself* (his Fig. 7, 1934).

This last discovery was one of the first indications of the tendency of galaxies to cluster, and was so noted by Hubble. In the 1934 paper he wrote:

[The log normal, rather than a straight N normal distribution is] "the feature [that] serves as a description and a measure of the tendency to cluster. It is clear that the groups and clusters are not superposed on a random (statistically uniform) distribution of isolated nebulae, but that the relation is organic."

While it is not clear what he meant by the last phrase of being organic, it is known (his comment once to me) that he knew that the distribution of the growth of bacteria in petri dishes in the laboratory show a log normal distribution of counts, and that after some time, clusters or colonies describe the mature distribution across the face of the dishes. (See Saslaw 1989; Saslaw and Hamilton 1984; Crane and Saslaw 1986; Coleman and Saslaw 1990; Karasev 1982, for modern discussions of the importance of a *log normal* rather than a direct normal distribution for the question of clustering).

Important as the 1934 paper was, no reliable apparent magnitudes could be attached to the counts. Indeed, the data were given as log N(E) (Hubble's Fig. 2), where E are the various exposure times of the photographic plates taken in the program.

As a final step, approximate conversion to magnitudes was then made by considering the reciprocity failure "Schwarzschild p exponent" in I$\sim$E^p for the intensity (I) and exposure time (E). In this way Hubble could assert that the galaxy counts continued to increase approximately as would be required as N(m)~0.6m if galaxies are distributed homogeneously in Euclidean space in the absence of all effects of redshift.

2.2. *The 1936 observational campaign*

In a most important paper two years later, Hubble (1936b) made an attempt to reduce the count data to a reliable system of magnitudes and to push the counts to a fainter limit.

It is important to point out that none of the photometry was done on individual galaxies as is done today. Rather, the "limiting" magnitude of plates taken with a particular exposure time, "reduced to standard conditions" ("full" photometric development, particular seeing conditions, particular emulsion batch, particular telescope, and standard photometric conditions) was estimated, based on comparisons using standard *stars*. The next step was to estimate the difference in the limiting magnitude between stars and the in-focus galaxy images. This was accomplished by a series of experiments (Hubble 1932, 1936b), among which were out-of-focus images of stars made to resemble galaxy images of particular sizes. In this way, the "limiting magnitude" of galaxies was estimated on the "standard condition plates." These are the magnitudes listed for each of the log N values in Table IV of Hubble (1936b).

There are two major problems with this procedure. (1) The apparent magnitudes of the stars used as standards had large systematic errors starting as bright as $m_{pg} = 16$, conclusively demonstrated only as late as 1950 (cf. Stebbins, Whitford, and Johnson 1950, see later). (2) Stars in the Mount Wilson Catalog used as standards reached magnitudes only as faint as $m_{pg} \sim 18.5$ in most of the Selected Areas, considerably brighter than what was needed by Hubble for his deepest counts. Hubble (1936b) writes "the estimation of the limiting magnitude for 2-hour exposures necessarily involved considerable extrapolation." The limit for his faintest counts was eventually listed as $m_{pg} = 21.03$.

Furthermore, even as Hubble's survey work was proceeding in 1934, Baade, whose main Mount Wilson duties were to determine scale errors in the Mount Wilson Catalog, was discovering substantial errors in Selected Area 68 which was the principal Area used by Hubble in his 1929 study of M31. The deviations from a Pogson scale began as bright as $m_{pg} \sim 17$. Baade's methods were still photographic, but now, using platinum neutral half filter methods (e.g., Weaver 1946; Stock and Williams 1962), his results were a substantial advance over the multiple exposure plus graded diaphragm methods used by Seares for the Mount Wilson Catalog† between 1910 and 1925.

Hubble's (1936b) final table of log N(m) values at faint magnitudes shows five data points for the counts at m_{pg} magnitudes of 18.47, 19.0, 19.4, 20.4, and 21.03, plotted as Fig. 1 here from Fig. 1 of Hubble. The analysis for the curvature of space and Hubble's answer whether the redshift is a true Friedmann-Lemaitre expansion depended on these five points.

Plotted are the integral counts as the log of the number of galaxies per square degree that are brighter than apparent magnitude m. The line labeled "Uniform Distribution" has a slope of dlog N(m)/dm = 0.6. The five points show a shallower slope. The

† Baade never published his new photometry in any complete detail, although he did summarize his corrections to Hubble's M31 magnitude scale in the paper announcing the resolution of the disk of M31 into stars (Baade 1944). Baade needed the faint magnitudes, transferred from his new scale in SA 68, to estimate that the resolved stars in the M31 disk had absolute magnitudes of $M_{pg} \sim -1.5$ and therefore that they are similar to globular cluster stars at the top of the giant branch. This connection played a central role in Baade's development of the population concept (Sandage 1986).

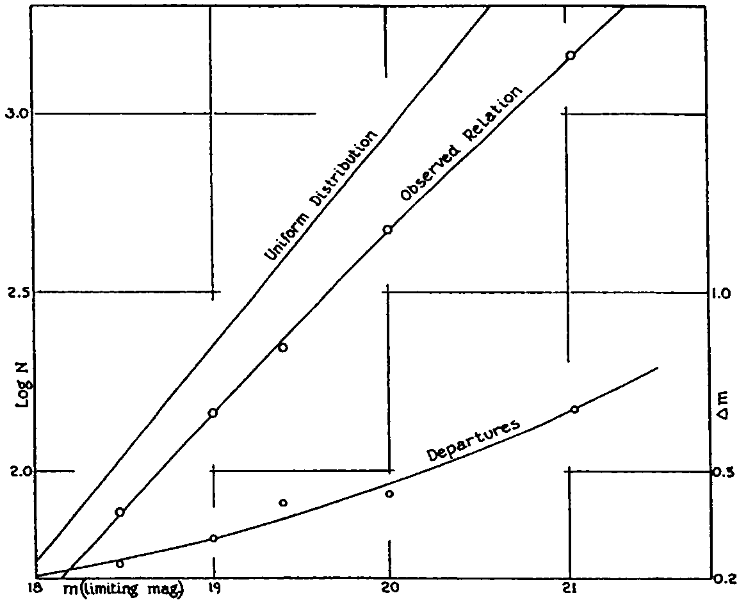

FIGURE 1. Hubble's final formulation of the log N(m) integral count-magnitude relation upon which his subsequent analysis of spatial curvature and the question of the reality of the expansion was based. N(m) is the number of galaxies per square degree brighter than apparent magnitude m. Diagram from Fig. 1 of Hubble (1936b).

departures of the five points from the "Uniform Distribution" line is shown as the lower curve. It is this "departure" curve that comprise the entire data set discussed by Hubble concerning the curvature of space and the reality of the expansion.

2.3. *Hubble's interpretation*

We now come to one of the most remarkable episodes in all of science. Hubble's (1936b) detailed analysis of Fig. 1 is a most fascinating study of how an interpretation, without caution concerning possible systematic errors, led to a conclusion that the systematic redshift effect is probably not due to a true Friedmann-Lemaitre expansion, but rather to an unknown, then as now, unidentified principle of nature. Indeed, even in the abstract to this 1936 paper on the "Effects of Redshift on the Distribution of Nebulae" Hubble concluded: "The high density suggests that the expanding models are a forced interpretation of the data." His belief that the expansion probably is not real persisted even into his final 1953 paper which was the Darwin lecture of the RAS, given in May of the year he died in September. What were the steps leading to this conclusion that, in today's climate, seems so remarkable?

Redshifts, ubiquitous for all galaxies everywhere, decrease the received flux. The larger the redshift, the larger is the decrease. This effect is one of the two principal reasons for the observed departure of the data in Fig. 1 from the "uniform distribution" supposition. The second is the departure of the intrinsic geometry from Euclidean, measured by "curvature" in the sense introduced by Gauss (1827, 1873). Hubble considered both of these effects.

There is no redshift information in Fig. 1. Hence, a relation between redshift and apparent magnitude is required to change the N(m) relation to the more fundamental N(z) function that is needed to make the calculations. And, because the counts in

Figure 1 are for field galaxies, the (m.z) relation for clusters as determined by Hubble and Humason (1931) and Humason (1936) had to be replaced by data for field galaxies. The required (m,z) relation was set out in Fig. 2 of Hubble (1936b), based on data of Hubble & Humason (1934).

Using the field galaxy (m,z) ridge-line relation of log cz = 0.2m + 0.77, the N(m) data of Fig. 1 could be changed to an N(z) relation, but only after correcting the observed magnitudes for the technical effects of redshifts, expressed as a K term, both selective and neutral (more later).

Armed now with a corrected N(z) relation, an assumption must be made concerning the relation between "distance" and redshift so that the N(z) relation could be changed to an N(r) relation, assumed to be proportional to the volume contained within "distance" r, leading to the geometry that is either Euclidean if (vol$\sim r^3$), or non-Euclidean if otherwise.

Note the extremely complicated multiple steps and, further, the questionable approximation of replacing distribution functions by mean values; viz. the counts in m are changed to N(z) and the K(z) term is changed to K(m) via an assumed *mean* (m,z) relation rather by using a luminosity distribution function for absolute magnitudes that takes into account the intrinsic spread in m at a given z.

A few of the intricate details of Hubble's procedure are set out in brief in the next section. Here we only summarize his conclusions, based on his analysis of the corrections to apparent magnitudes due to the effects of redshifts.

As is now well known, if redshifts are due to a true expansion, the required correction to the observed apparent magnitudes are by two factors of (1 + z) for the so called number effect (the number of photons received per second from a receding source) plus the energy effect [each photon is degraded in energy by the redshift, again by (1 + z)]. Hubble concluded (see the next section) that if two factors of (1 + z) are applied to his Fig. 1 data, then the curvature correction needed to make the data conform to the "Uniform Distribution" condition would have to be enormous, giving a very small, high density, large curvature universe, so small and of such high density that Hubble believed that the procedure gave impossible results. He continued to write his conclusion to the end, calling into question the reality of the expansion that required the second factor of (1 + z) correction for the "number effect."

To make understandable the language of Hubble's analysis, the K correction as used by Hubble is the selective effect (plus the bandwidth effect)† of shifting a galaxy spectrum through the blue photographic band pass "plus" either the one or two factors of 2.5 log (1 + z). Hubble expressed the total effect as $K = B \times z$.

Hubble determined B from the observations using the "departure" curve in Fig. 1. His program was then to compare this B(observed) with a theoretical B^* calculated using the assumption of either a true expansion or not (either one or two factors of 1 + z), plus the selective term found by shifting an assumed galaxy spectrum through the photometric band pass (the selective part of the K term; see Humason et al. 1956, Appendix B; Oke and Sandage 1968). In what follows, the argument hinges on the comparison of B and B^*.

† The spectrum is stretched by multiplying each rest wavelength by a factor of 1 + z. By so doing, compensation must be made for the decreased bandwidth of the stretched spectrum over the fixed sensitivity pattern of the plate, filter, and telescope photometric functions. I mistakenly have written (Sandage 1995, Lecture 2) that Hubble neglected the bandwidth correction. However, a detailed examination of his calculated K corrections for a blackbody with T = 6000° K (Hubble 1936b, his Tables V and VI) shows that he did not neglect the bandwidth term and that my comment in the Saas Fee Lectures is incorrect.

Hubble's analysis (page 533 of the 1936b paper) of his "departure" curve gave B = 2.94. His *calculated* K term (assuming a black body galaxy spectrum of T = 6000° K) was either $B^* = 3.0$ for no expansion, or $B^* = 4.0$ for a real expansion (energy plus number effect). Clearly, only the no-expansion solution fitted Hubble's putative B = 2.94 departure data in Fig. 1.

Many conclusions were made from this result, not only concerning the reality of the expansion but also concerning the consequences of a real expansion for a second-order term in the velocity-distance relation as the measurement of deceleration, the value of the space curvature, and the question of evolution of galaxy absolute magnitudes in the look-back time. Several of the direct quotes concerning these issues are of interest for the work of the present workshop.

On his page 542: "It is evident that the observed result, B = 2.94, is accounted for if redshifts are not velocity shifts. The comparison is based on an effective temperature, T_o of 6000°, but the uncertainties cover the range down to about $T_o = 5750$. The interpretation is consistent with the data [but only if]—the *expansion and spatial curvature are either negligible or zero*." (Emphasis added).

Concerning the redshift-distance relation; page 38 of Hubble (1937): "The inclusion of recession factors would displace all the points [in the Hubble diagram of redshift vs. apparent magnitude of great clusters—his Fig. 1 of the 1937 reference] to the left [higher redshifts at a given magnitude], thus destroying the linearity of the law of redshifts". [N.b., this is not correct when the appropriate Mattig relations for the (m,z) Hubble diagram are used; see later].

For the conclusion on the reality of the expansion (Hubble 1936b, page 553):—"if redshifts are not primarily due to velocity shifts, the observable region loses much of its significance. The velocity-distance relation is linear; the distribution of nebulae is uniform; there is no evidence of expansion, no trace of curvature, no restriction of the time scale." Page 553/4: "The unexpected and truly remarkable features are introduced by the additional assumption that redshifts measure recession. The velocity-distance relation deviates from linearity by the exact amount of the postulated recession. The distribution departs from uniformity by the exact amount of the recession. The departures are compensated by curvature which is the exact equivalent of the recession. Unless the coincidences are evidence of an underlying necessary relation between the various factors, they detract materially from the plausibility of the interpretation.—the small scale of the expanding model, both in space and time is a novelty, and as such will require rather decisive evidence for its acceptance."

From his Darwin lecture (Hubble 1953): "When no recession factors are included, the law will represent approximately a linear relation between red-shifts and distance. When recession factors are included, the distance relation is expected to be—non-linear in the sense of accelerated expansion" [sic, not the correct sign; the word must clearly be *decelerated*, as he in fact wrote twice earlier in 1936 and 1937] . "—[If no recession factor is included] the 'age of the universe' is likely to be between 3000 and 4000 million years, and thus [again with no recession factor] comparable with the age of rock in the crust of the Earth."

Concerning the second-order term in the velocity-distance relation: (1936b), page 546: "Since the second-order term [with recession factors included] is definitely positive, the possible models are restricted to those in which the rate of expansion has been diminishing during the past several hundred million years." And again in (1937, page 43): "The chief significance of the term for cosmological theory lies in the positive sign [of the redshift vs. distance correlation]. The rate of expansion of the universe has been slowing down, at least for the past several hundred million years. The 'age of the universe' is considerably shorter than that permitted by the linear law."

Finally, as to evolution of luminosity in the look-back time (Hubble 1936b, page 543). "As for the constancy of nebular luminosities, the question is whether or not luminosities of spirals change materially (say 10%, or 0.1 mag) during [the look-back time]. —very few students will hesitate to adopt the assumption that systematic variation in so short a [time] interval will be inappreciable." Reasons are then given in the remainder of the paragraph, none of which would pass today's referees armed with the present knowledge of population synthesis and stellar evolution.

The clearest proof that Hubble maintained these views concerning the reality of the expansion until the end is the style of argument in his 1953 Darwin Lecture, seen in particular from Fig. 1 of that lecture. No recession factor was put to the K-correction magnitudes in the abscissa of the m,z (Hubble) diagram. This diagram was the first rendering in the literature of that central cornerstone of observational cosmology using new data from the Palomar 200-inch reflector.

The several arguments set out above were of particular importance for the Palomar program on these matters that followed from 1953 through the 1980s. Reasons why Hubble's conclusions should be changed emerged slowly from this program, not only because of new observations based on photoelectric photometry, but also from a much deeper understanding of the interface between the observations and the theory (section 4).

3. Why Hubble's program failed

What then was wrong with Hubble's 1936 analysis of the count data in Fig. 1 that led him to his remarkable conclusion of no expansion?

There were five problems. (1) Incorrect K term values as a function of redshift because galaxy spectra have a much cooler color temperature than 6000°. (2) The apparent magnitude scale used by Hubble via the Selected Area magnitudes, even as partially corrected by Baade in the late 1930s, was wrong. (3) Hubble's assumption that "distance" is given by cz/H for large redshifts is not correct, but known only after the Mattig revolution (section 4). (4) The assumption that uniform spatial distribution requires log N(m) to increase as $\sim$0.6 m for large redshifts is also wrong according to the theory of Friedmann spaces, again shown by the new Mattig equations. (5) The assumption of constant luminosity and/or density evolution at high redshifts is evidently wrong as shown by the large excess in the counts (a fact that would have been discovered by Hubble from his counts if he had kept the true expansion assumption) shown not only by the modern N(m) counts, but also by the strange galaxy morphology at the faintest HST levels (section 7).

These points are reviewed in order.

3.1. *Enter Greenstein*

The 1936 analysis by Hubble had already begun to unravel by a devastating paper by Greenstein (1938), in which he showed that the color temperature of M31 was only 4200° K rather than 6000°. Shifting a black body spectrum through the m_{pg} pass bands gave selective K corrections plus *either* one or two factors of 2.5 log (1 + z) that were no where near the B = 2.94 determined from the "departure" observations by Hubble. Hence, nothing worked in *any* interpretation of Hubble's 1936b counts. Greenstein's conclusion was: "From λ6500 to 3900 Å, [the spectrum of M31] closely resembles that of a black body of temperature 4200°.—The effect of such low temperatures on the present interpretation of counts of extragalactic nebulae is serious. It seems improbable that the effect of the redshift on the apparent magnitudes of nebulae, found by Hubble, can be interpreted either as a velocity or as a nonvelocity shift."

3.2. *Modern K term*

However, the problem with the K-term was even more serious because it was soon realized, via the non-existent Stebbins-Whitford (1948) effect and its explanation (Oke and Sandage 1968), that galaxy spectral energy distributions (SED) are very poorly approximated as black bodies, primarily because of the important 4000 Å break. Improvement of the SEDs that had been measured by Oke and Sandage was made by Whitford (1971). Oke and Sandage had observed only the central regions of five giant E galaxies using a spectrum scanner at the Cassegrain focui of both the Mount Wilson 60 and 100-inch reflectors. Whitford could observe a much larger fraction of the total E-galaxy light with his spectrum scanner at the Lick Crossley reflector because its shorter focal length and therefore smaller focal plane scale. The radial color gradient of E galaxies, becoming bluer from the center to the outside, explained the 10% difference between the two studies.

The resulting mean SEDs, shown in Fig. 2, permitted, for the first time, the calculation of realistic K(z) corrections for the effects of shifting the mean E-galaxy SED through various photometric pass bands. Since then, a large industry of K term calculations has developed, not only for E galaxies but for all Hubble types. A more modern history with entrance to the extensive literature is given elsewhere (Sandage 1988, 1995).

The result of Whitford's calculations (his Table 3), approximated as the first term of a power series in z, is

$K_B = 7.1z$ if no recession (only the energy effect term), and

$K_B = 8.1z$ if the number effect term for real recession is included.

Clearly, neither of these cases are consistent with Hubble's putative 1936 requirement that $B_{mpg} = 2.94$.

3.3. *Corrections to the magnitude scale*

As mentioned earlier, in one of the most important papers of the decade, Stebbins, Whitford, and Johnson (1950) showed that corrections to the m_{pg} magnitudes in the Mount Wilson Catalog for SA 57, SA 61, and SA 68 begin as bright as $m_{pg} = 16$ and increase with increasing faintness. They further showed that the North Polar Sequence, long the principal source of faint standards in the decades before 1950 (Seares 1915, 1922a,b; Seares and Humason 1922) was accurate in scale to better than the 0.1 mag level to the limit of its tabulation.

Later work extended the result to additional selected areas where it was shown that the needed corrections were generally ubiquitous and in some areas reached 1.5 mag at m_{pg}(Seares) = 18.5. Examples for four Selected Areas are in Figure 3, taken from an early summary of a systematic program to determine the corrections to the faintest level of the Catalog (Sandage 1983, 1998) in a dozen Selected Areas. The corrections begin near $m_{pg} = 15$ and increase to an average of 0.7 mag at $m_{pg} = 18$.

4. The Mattig revolution 1958–1959

Now we come to the heart of the problem with Hubble's interpretation, but more importantly to the watershed for practical cosmology itself in a fundamental development that changed the field.

Reading most of the many papers on observational cosmology before the early 1960s, nowhere does one see the modern approach of solving the Friedmann equation that

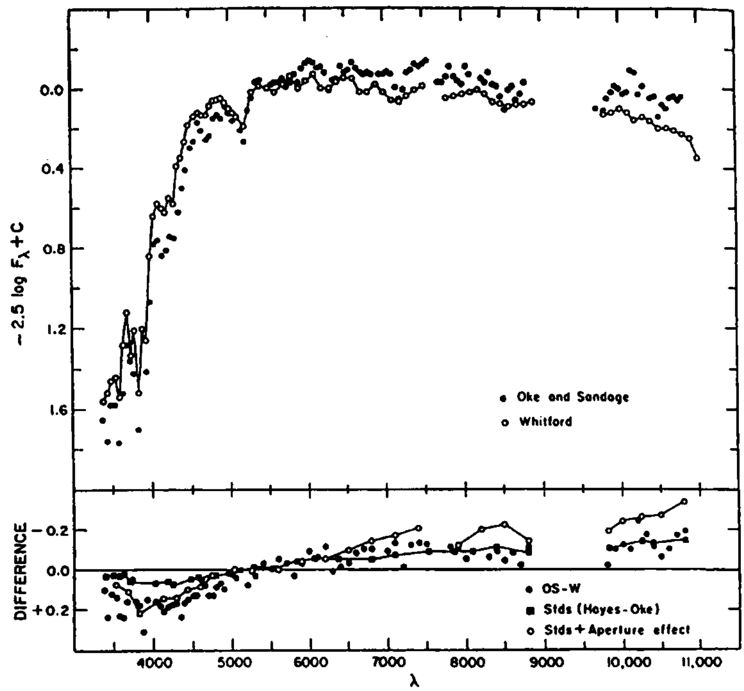

FIGURE 2. Mean spectral energy distribution of giant E galaxies as measured by Oke and Sandage (1968) for the very central regions and by Whitford (1971) for larger scanning apertures. These are the SEDs used for the first reliable calculation of the selective part of the K corrections in B, V, and R pass bands for E galaxies. The calculated K term in B using these SRDs was very much larger than the m_{pg} K correction used by Hubble (1936b). Diagram from Whitford (1971).

describes the development of the scale factor R(t) with time, and how the closed form of R(t) for arbitrarily high z is to be used to obtain the exact equations necessary to interpret the observations, valid for all redshifts. What in fact is the correct equation for the interval "distance," r, as a function of redshift?

Before the correct equations became known in the late 1950s, all relations involving observed magnitudes, angular diameters, redshifts, spatial volumes, and the consequently N(m) counts were given in Taylor series expansions in z, using only $R(t_o)$ and the first several derivatives of R(t) about the present epoch. The only assumption on R(t) was that it is a smooth enough function for the Taylor expansion to mean something. These series expansions, while good for small redshifts, were worthless for redshifts larger than perhaps 0.3, which was in fact near the limit of redshifts known even in the late 1950s.

What is remarkable about this situation is that the Friedmann equation and its solution never entered most of these papers at the interface between theory and observation. Examples are the marvelously complicated series-expansion papers by Davidson (1959a,b, 1960), and even more remarkably, the first edition of the famous text book by McVittie (1956).

The developments that began the modern era were the derivations of all the relevant equations using the Friedmann equation to give the explicit solutions of R(t) for all t (i.e., for arbitrarily high redshifts). Remarkably, the complete development is set out in two short papers by Mattig (1958, 1959).