The 'fitting problem' in cosmology



1987 Class. Quantum Grav. 4 1697

(http://iopscience.iop.org/0264-9381/4/6/025)

Class. Quantum Grav. 4 (1987) 1697-1729. Printed in the UK

# The 'fitting problem' in cosmology

G F R Ellis† and W Stoeger‡

† School of Mathematics, Queen Mary College, Mile End Road, London E1 4NS, UK and Department of Applied Mathematics, University of Cape Town, Rondebosch 7700, South Africa

‡ Vatican Observatory, Castel Gandolfo, I-00120 Citta del Vaticano

Received 6 February 1987

**Abstract.** This paper considers the best way to fit an idealised exactly homogeneous and isotropic universe model to a realistic ('lumpy') universe; whether made explicit or not, some such approach of necessity underlies the use of the standard Robertson-Walker models as models of the real universe. Approaches based on averaging, normal coordinates and null data are presented, the latter offering the best opportunity to relate the fitting procedure to data obtainable by astronomical observations.

## 1. Introduction

Modern cosmology aims at finding the large-scale matter distribution and spacetime structure of the universe from astronomical observations. There are broadly speaking two distinct approaches that have been applied to this problem.

### *1.1. Approaches to observational cosmology*

The standard approach is to make some *a priori* assumption about the geometry of the universe, based either on philosophical or pragmatic grounds; usually this assumption is that the universe is spatially homogeneous and isotropic on a large scale (this is the *cosmological principle*; Bondi (1960), Weinberg (1972)). It then follows that the universe is accurately represented by a Friedmann-Lemáitre-Robertson-Walker (FLRW) model, and the primary objective of observational cosmology is to determine the two or three free parameters characteristic of such universe models (Sandage 1968).

An alternative is to attempt as far as possible to determine spacetime geometry directly from astronomical observations without making such *a priori* assumptions (Kristian and Sachs 1966, Ellis *et al* 1985). However difficulties arise in actually determining the geometry of spacetime in this way when realistic observational limitations are taken into account (Ellis 1980).

These approaches are both to some extent unsatisfactory—the first because the universe is manifestly *not* a FLRW universe on at least some scales, and the approach does not provide any guidance as to on what scales such models are supposed to be applicable, nor seriously consider the issues arising when we consider the relation between descriptions of the universe at different scales of inhomogeneity (Ellis 1984); and the second because of the practical difficulties in implementation (theoretically determinable quantities are very difficult to determine in practice). Alternative analyses

0264-9381/87/061697+33$02.50  

of homogeneity based on 'almost Killing vectors' (Matzner 1968) or 'observational homogeneity' (Bonnor and Ellis 1986) do not yet seem able to resolve these issues.

### *1.2. The fitting problem*

A further possible approach, to some extent intermediate between the two outlined above, is based on what may be called the 'fitting problem' (Ellis 1984). The basic idea here is that we do not *a priori* assume the universe is necessarily well described by a FLRW model at all times, but nevertheless decide to use such a model for pragmatic and historical reasons. The issue facing us then is as follows: we contemplate on the one hand, a ('lumpy') cosmological model $U=\{M, g_{ab}, u^a, \mu, n\}$ comprising a manifold M, metric tensor $g_{ab}(x^j)$, normalised 4-velocity $u^a(x^i)$: $u^a g_{ab} u^b=-1$ (see, e.g., Ellis 1971), dynamical matter variables symbolised here by the energy density $\mu$ (but in general including other quantities such as the pressure $p$), and other matter variables symbolised here by the galaxy number density $n$ (but in general including more detailed specification of the distribution of luminous matter in the universe), which together give a realistic representation of the universe including all inhomogeneities down to some specified length scale $L$; and on the other an idealised, completely smoothed-out FLRW model $U'=\{M', g'_{ab}, u'^a, \mu', n'\}$. Our problem is how to determine a 'best fit' between these two cosmological models (figure 1($a$)). This will then determine the appropriate way we should use the idealised FLRW cosmology $U'$ to model the irregular nature of the real universe (represented here by the more realistic 'lumpy' universe model U). The approach resembles that used in geodesy, where a perfect sphere is fitted to the 'pear-shaped' earth; deviations of the real earth from the idealised model can then be measured and characterised (King-Hele and Walker 1986, figure 1($b$)).

This way of approaching the issue of observations in cosmology has the potential to throw light on the geometrical and physical interpretation of the FLRW models we use, precisely because it focuses on the relation between the idealised smooth model and more realistic descriptions of reality. In particular it should enable one not only to determine a best fitting FLRW universe model, but also to specify details of that fitting (giving the difference between the two models at each spacetime point), and thus to investigate how good the fit is (characterising how adequately the realistic model is approximated by the smoothed-out model). Part of the utility of the approach is that if a suitable procedure is implemented, it might be possible to use it repeatedly; that is, to consider a best fit between any lumpy universe model $U'$ and a model $U''$ that gives an even better description of the real universe than $U'$ by describing the inhomogeneities at an even more detailed level. This process would in principle allow one to determine the best description at any prescribed level of detail.

There are many ways this approach might be tried; implementations could be based on (i) the space of spacetimes; (ii) initial data for spacetimes; (iii) the 'gauges' adopted in perturbation studies; (iv) near equivalence; (v) average behaviour; (vi) null data; and (vii) normal coordinates. We have considered each of these possibilities and the desirable conditions to be satisfied by a useful fitting procedure. This has led us to conclude that while each of these ways of tackling the problem is potentially interesting, the null data approach seems to have advantages over the others particularly in terms of practical implementation. Nevertheless all are worth consideration; we now outline the last three approaches, each of which brings out an important aspect of the fitting problem.

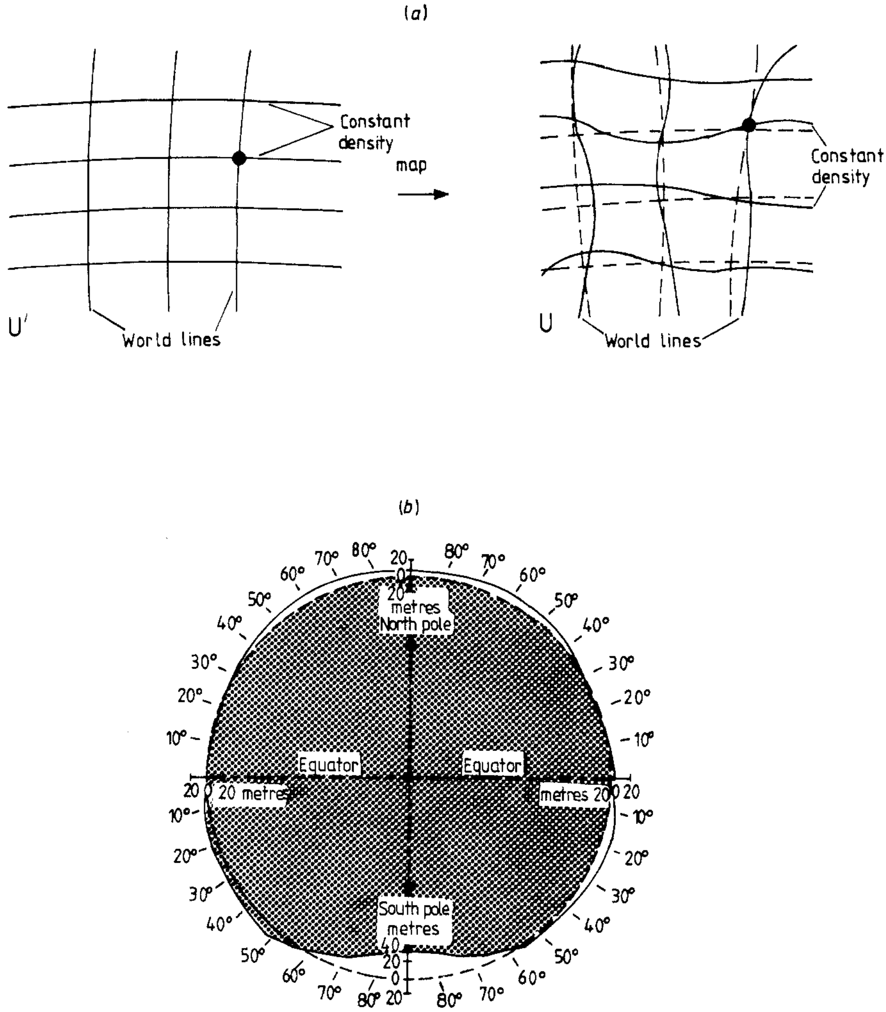


**Figure 1.** (*a*) An exactly uniform and spherically symmetrical FLRW universe U′ mapped into the lumpy universe U so as to give the best fit possible. (*b*) An exactly spherical sphere fitted to the lumpy world to give the best fit possible.

### 1.3. *Averaging*

One important way of thinking of the use of a smoothed-out model is that it represents the average properties of the lumpy model. Let the smoothed-out description U* be obtained from a lumpy spacetime U by a suitable averaging procedure; it then represents the nature of U when described over some averaging scale $L$. If indeed it is sensible to describe the large-scale nature of U by some Robertson–Walker spacetime, then the best-fit FLRW model U′ should be the same as the averaged model U*. A suitable fitting procedure would hopefully include this way of looking at the relation between U and U′. In particular it should determine the appropriate averaging scale associated with the smoothed-out model: that is, it should lead to a statement that the universe U can reasonably be regarded as a FLRW universe U′ *if* averaged out over a specified length scale $L$.

### *1.4. Fitting on the basis of null data*

A meaningful 'best-fitting' procedure in cosmology should be related directly to possible astronomical observations. We can go to the heart of the issue by focusing on fitting the null data for cosmology on the past light cones $C^-(p)$ and $C^-(p')$ of points p, p′ in U, U′ respectively. Further advantages of this approach are that most of the arbitrariness of choice of the initial data surface that plague other approaches disappears in this case, there are no constraints for initial data on characteristic surfaces, and the causal nature of what is and is not possible in observational cosmology (cf Ellis 1975, 1984) is necessarily brought to the fore by this approach.

Such a fitting procedure has several parts (figure 2). (*a*) Choose the vertex points p, p′ to compare in U, U′. (*b*) Fit 4-velocities $u^a$, $u'^a$ in the two models at the points p, p′. (*c*) Isotropise on the past null cones $C^-(p)$, $C^-(p')$ to obtain averaged (spherically symmetric) geometries. This primarily involves choosing a measure of distance down the null cone that will distinguish particular 2-spheres to use in this averaging procedure. (*d*) Compare these spherically symmetric geometries in depth down the two light cones, finding a 'best-fit' FLRW universe and a best-fitting of that universe. This provides the basis for tests of the degree to which the lumpy universe U is approximated by the FLRW model U′ on the past light cone. (*e*) The fitting has now taken place on the past null cone. Finally construct the best-fit spacetime itself inside and outside the null cone from the initial data on the null cone, and, insofar as is possible, estimate how good the fit is off the light cone (problems arise outside the past null cone, cf Ellis (1984)).

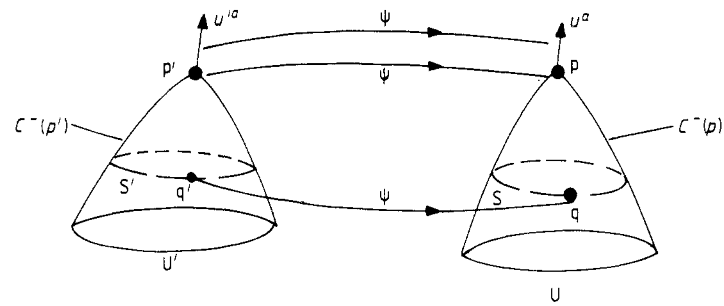


**Figure 2.** In the null data best-fitting, one successively chooses maps from U to U′ of the null cone vertex p′, the matter 4-velocity at p′, a 2-sphere S′ on the null cone of p′ and a point q′ on the 2-sphere. This establishes the correspondence $\psi$ of points on the past null cone of p′ to the past null cone of p, and one then compares initial data at q′ and at q.

### *1.5. Local physics and normal coordinates*

As well as being fundamental to observational cosmology, the question of fitting arises in investigating the effect of cosmological boundary conditions on local physics. For example it plays an important role in galaxy formation studies, where the question of correspondence between the smooth and lumpy universe models is essentially the same as the question of choice of gauge (Bardeen 1980). One can make some progress without specifying the fitting if one works in terms of gauge-invariant quantities (Bardeen 1980) but a linearised quantity is gauge invariant only if its unperturbed value has a vanishing Lie derivative for all vector fields (Sachs and Wolfe 1967, Stewart and Walker 1974). Higher-order terms are needed to obtain an approximation in which gravitational self-interaction effects are taken into account, and already at second order there are no gauge invariant quantities. Thus even for the Newtonian limit in which self-gravitation is included—which must surely be needed to study galaxy formation problems properly—the answers obtained necessarily depend on the correspondence chosen.

This implies that having chosen a fitting procedure based on observations, it is important to ask what this implies about local physics in the expanding universe. Normal coordinates can be thought of as fitting the lumpy universe to flat spacetime in a small neighbourhood of the origin, where (by the equivalence principle) curved spacetime physics must be nearly but not exactly the same as flat spacetime physics. Thus the use of normal coordinates is one way to look at the relation between cosmology and local physics; and one can in particular examine the implications of a fitting procedure for local physics by comparing the use of normal coordinates in the lumpy universe U and the smoothed-out model U′.

### *1.6. This paper*

We first consider the issue of averaging a lumpy universe to obtain a smooth model in § 2. Then after summarising the nature of observational coordinates and observational quantities in cosmology in general, and in FLRW universes in particular, in § 3 we examine the null data approach to the fitting problem in § 4. It will turn out that in practice such an approach is similar to what is done at present by observers, but with some potentially important further features; in particular, it can be related to the averaging approach, and suggests the nature of possible criteria of acceptable fit, determining if in fact it is reasonable to use U′ to represent U at the chosen scales. The importance of having fitting criteria is underlined by a brief examination of the normal-coordinate approach to optimal fitting in § 5, which provides a view of the interaction of cosmological conditions and local physics. In order to evaluate the proposals made here it is important to consider the other possibilities, so alternative best-fitting proposals are presented in the appendix.

We believe the suggestion made here is a useful approach to making explicit the way FLRW models are used as a basis for representing nearly-FLRW spacetimes. This is a fundamental issue in cosmology, both in terms of interpreting the meaning of FLRW models proposed as representing the actual universe, and in relation to galaxy formation studies. The relation between the fitting procedure ('choice of gauge') proposed here and perturbation (galaxy formation) studies will be pursued in a subsequent paper. Even if the particular approach suggested here ultimately turns out to be unsatisfactory, the paper will have served a useful purpose if it brings to wider attention the issue of the 'fitting problem' in cosmology, which is of importance and deserves considerable thought.

## **2. Average data**

This approach to fitting a FLRW model U′ to a 'lumpy' model U, following the work of Carfora and Marzuoli (1984), is based on the concept that the smoothed-out model

U′ should accurately represent the average behaviour of the more realistic model U (cf § 1.3).

### 2.1. Compact spatial sections

This is easiest to consider in the case where the spacetimes are each the direct product of closed (compact) space-sections $\Sigma$, $\Sigma'$ and a time parameter $t$. This requires that the smoothed-out FLRW model either has positive spatial curvature ($k=+1$), or is a 'small universe' with unusual topology (Ellis and Schreiber 1986). If we demand that the smoothed-out model U′ should represent the average behaviour of the 'lumpy' model U, what specific requirements does this imply? To compare them we need to choose an appropriate proper-time labelling $\tau$, $\tau'$ for the families of surfaces $\Sigma$, $\Sigma'$ in the two universes. We then might specify the FLRW model by the requirements that

(i) the volume behaviour of the two universes be the same, i.e. the total volume $V$ of the space sections $\Sigma$ is to be the same function of $\tau$ as the total volume $V'$ of $\Sigma'$ is of $\tau'$;

(ii) the average energy density $\mu'$ on $\Sigma'(\tau')$ be the average over $\Sigma(\tau)$ of $\mu$ for the value $\tau$ corresponding to $\tau'$, and the average pressure $p'$ on $\Sigma(\tau')$ be the average over $\Sigma(\tau)$ of $p$. In this case the smoothed model U′ would precisely reproduce the average volume, density and pressure behaviour of the more realistic lumpy universe model U.

The first problem is that while there is an obvious unique choice of spatial slicing in the FLRW case (universe U′), there will in general be no unique choice of such slicing in U unless there are surfaces of homogeneity in U. Thus the criterion is not unique: we may get conflicting answers by taking different slicings $\Sigma(\tau)$ of the same universe U. We might for example choose (*a*) surfaces of constant density, (*b*) surfaces orthogonal to the fluid flow in the case of vanishing vorticity, or (*c*) surfaces of minimal distortion†, but the appropriateness of these choices would need investigation. This problem is exacerbated because even given a choice of these surfaces, the proper time labelling $\tau$ of $\Sigma(\tau)$ will in general not be unique. The difficulties here are essential—they will arise in any attempt to discuss the time evolution of the universe by an averaging process defined on spacelike hypersurfaces.

Secondly, the averaging of volumes, energy and pressure will in general lead to a FLRW model which does not obey the Einstein field equations, because the processes of spatial averaging and of going to the Einstein field equations do not commute (Ellis 1984). More specifically, if we substitute the radius function $R'(\tau')$ corresponding to the average volume behaviour $V'(\tau')$ into the Einstein field equations for the FLRW universe U′, we will derive from them an energy density $\mu^*(T)$ and pressure $p^*(\tau)$ that are *not* the same as $\mu'(\tau)$ and $p'(\tau)$. The way round this is to regard $\mu^*$ and $p^*$ (which are the *effective* values of the energy density and pressure) as containing contributions from the smoothing out of the lumpiness at more detailed scales of description (Carfora and Marzuoli 1984); physically these represent the contributions to the effective overall energy density from gravitational waves, binding energy of compact objects, etc (which are not included in the averaging procedure leading to $\mu'$ and $p'$). We define the averaged FLRW model simply by its volume behaviour; the difference between the averaged density $\mu'$ and pressure $p'$ and the implied FLRW values $\mu^*$ and $p^*$ is then defined to be the contribution to these quantities due to small-scale inhomogeneity.

† In a FLRW or almost FLRW universe there will in general not be any hypersurfaces of maximal volume.

Given this understanding, the averaging procedure can indeed provide a suitable fitting procedure in that, given the choice of spacelike slicing, a unique FLRW model U′ is associated with U. However problems remain apart from the arbitrariness of choice of spatial slicing and of proper time $\tau$; for example, the relation between astronomical observations in the two models is obscure. Nevertheless the concept of averaging seems to be fundamental to the interpretation of the smoothed-out models, so this procedure deserves further investigation.

### *2.2. Non-compact spatial sections*

Can one hope to use a similar method in the cases where $\Sigma$, $\Sigma'$ are infinite (as will follow in the FLRW $k=0$ and $k=-1$ cases with the usual topology)? Our problem now is that even given the spatial sections $\Sigma$, there is no longer defined a unique finite volume over which to take averages. The remedy is simply to *choose* $(a)$ an initial surface $\Sigma_0$ within the family of surfaces, and $(b)$ within this a suitable averaging volume with an arbitrary centre and specified length scale. Thus we take averages over a volume $V_L(x)$ of length scale $L$ in the initial hypersurface $\Sigma_0$, centred at a point $x$ in $\Sigma_0$. At earlier and later times, we define the corresponding volume $V_{L,x}(\tau)$ in $\Sigma(\tau)$ by following the fluid flow, that is $V_{L,x}(\tau)$ is comoving with the fluid and coincides with $V_L(x)$ on $\Sigma_0$. We can then take averages as before on these volumes $V_{L,x}(\tau)$, determining averaged volume behaviour $V(\tau)$, energy density $\mu(\tau)$ and pressure $p(\tau)$. The same remarks apply here as in the compact case, but now there is an additional major factor: one can ask how uniform the result of the procedure is as we vary $x$ and $L$. The mapping defined from U to U′ will vary pointwise with $x$ and also with the scale $L$; that is, we obtain a specific FLRW model U′$(x, L)$ for each choice of $x$ and $L$. At first this seems like a major disadvantage, but in fact it is a plus: for this gives a criterion of acceptable fit. We can regard the lumpy model U as reasonably represented by the smoothed-out model U′ if U′$(x, L)$ is 'close' to U′, in some suitable measure, for all $x$ in $\Sigma_0$ when $L$ lies in some specified range of lengths $(L_1, L_2)$. If no model U′ can be chosen to satisfy this requirement, then U is not reasonably modelled by any FLRW universe model (cf e.g. Stoeger *et al* (1987) for a necessary condition on the density at each time $\tau$ if there is to be a reasonable fit). Note that the best-fit procedure and measure of goodness of fit required here can be quite simple, e.g. a $\chi^2$ test based on Hubble constant and deceleration parameter measurements, because they will simply be comparing between FLRW universes.

Such criteria for 'Robertson–Walker like character' of a universe model are one of the features we wish to obtain. Thus even in the compact case, when we can define an averaging over the whole spatial section, we might still find it useful to consider averaging the behaviour over smaller averaging volumes as in the non-compact case just mentioned. This procedure can help establish the degree of approximation by FLRW universe models to local regions of the universe; such criteria will be useful if they can be related to local astronomical observations. Of course it may easily be true that a FLRW model U′ will give a good fit only if, as well as restricting the averaging length scale $L$, we restrict the time of comparison $\tau$ to some limited range.

The problem with this approach, apart from the arbitrariness associated with the choice of the surfaces of constant time, is that it is not directly related to possible observations except on scales that are rather small (by cosmological standards). What one needs is some adaptation of this idea to observations obtainable on our past light cone. Thus we now turn to considering light cone observations and fitting based on them, and then return to considering averaging procedures in that context.

## 3. Null data for cosmology

The null data for cosmology and their relation to possible observations is described in Ellis *et al* (1985), which extends the power series analysis by Kristian and Sachs (1966).

### *3.1. Observational coordinates and the metric tensor*

The analysis is based on observational coordinates $(w, y, \theta, \phi)$, where $w$ labels the past null cones of a world line C, which in the situation we have in mind will be geodesic; $y$ is a measure of distance down these null cones (an affine parameter, cosmological redshift, area distance, or other suitable coordinate); and $\theta$, $\phi$ are standard polar coordinates based on a tetrad parallely propagated along C (figure 3).

When such coordinates are used, the metric tensor takes the following form:

$$g_{ab} = \begin{pmatrix} \alpha & \beta & \gamma_2 & \gamma_3 \\ \beta & 0 & 0 & 0 \\ \gamma_2 & 0 & h_{22} & h_{23} \\ \gamma_3 & 0 & h_{23} & h_{33} \end{pmatrix} \qquad (1)$$

where each of the metric components $g_{ab}$ are in general arbitrary functions of the coordinates $x^0 = w$, $x^1 = y$, $x^2 = \theta$ and $x^3 = \phi$. These components are required to obey suitable central conditions (Ellis *et al* 1985, appendix A). The coordinate $y$ can always be chosen to be comoving with the fluid (i.e. $y_{,a}u^a = 0$) but in general this is not possible for $\theta$, $\phi$ (because the angular coordinates are based on a non-rotating tetrad, and the average motion of matter may rotate relative to such a basis).

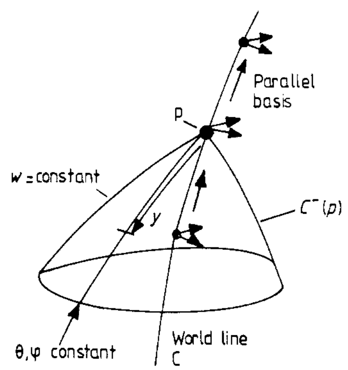


**Figure 3.** Observational coordinates on the past null cone of p. A parallel basis along the geodesic world line C forms a basis for angular coordinates $\theta$, $\phi$. The past null cones of points on C are labelled by the coordinate $w$, and $y$ is a measure of distance down the null cone from its vertex.

### 3.2. *Observable quantities*

Quantities in principle observable by astronomical observations are the redshift $z$, area distance $r_0$ (determinable in principle from angular diameter or luminosity measurements), energy density $\mu$ (determinable dynamically from observations such as galactic rotation curves), number counts $dN/dz$ in solid angle $d\Omega$ and background radiation temperature $T$. These quantities enable us to directly determine three components of the matter 4-velocity and three components of the metric tensor on the past null cone $C^-(p)$ of a point p on C, and also allow estimation of the matter content (numbers of objects, energy density of matter present). Given these data one can in principle use Einstein's field equations to determine spacetime completely inside $C^-(p)$, and hope to estimate its properties outside (§§ 12 and 13 of Ellis *et al* 1985).

In detail, the observable quantities on the null cone $w = w_0$ are as follows.

(i) Choosing $w$ to measure proper time along the world line C, the component $u^0$ of the matter 4-velocity is directly given by the observed cosmological redshift $z$:

$$u^0 = 1 + z. \tag{2}$$

(ii) Observed galactic proper motions $M^A(z, \theta, \phi)$ determine the angular components $u^A$ of the 4-velocity ($A = 1, 2$) through the relation

$$u^A = (1+z)^{-1} M^A. \tag{3}$$

(iii) Observations of an extended object of known size and shape can in principle directly determine the metric tensor angular components $g_{AB}(z, \theta, \phi)$ ($A, B = 2, 3$). It is convenient to separate out the area distance $r_0(z, \theta, \phi)$, defined by the relation

$$(r_0)^2 \sin\theta = (\det g_{AB})^{1/2}. \tag{4}$$

This quantity can in principle be determined either by area measurements or from observed luminosities (Ellis 1971). Then the rest of the information in $g_{AB}$ can be represented by

$$f_{AB} = \frac{g_{AB}}{(r_0)^2 \sin\theta} - f_{0\,AB} \tag{5}$$

where $f_{0\,AB}$ is the metric of the unit sphere divided by $\sin\theta$ (i.e. $f_{0\,22} = 1/\sin\theta$, $f_{0\,33} = \sin\theta$, $f_{0\,23} = f_{0\,32} = 0$); $f_{AB}$ is just the distortion in the image due to anisotropic gravitational lensing.

(iv) If we can estimate the masses of observed objects we can deduce from the number counts the effective energy density $\mu_{\text{lum}}$ due to this observed matter. In detail, the number density $n(z, \theta, \phi)$ of objects of given type at points on the past light cone is

$$n = N_0(z, \theta, \phi)(dz/dy)(1+z)^{-1} \tag{6a}$$

where

$$N_0(z, \theta, \phi) = \frac{1}{F\, d\Omega\, r_0^2} \frac{dN}{dz} \tag{6b}$$

can be determined from differential number counts $dN/dz$ in a solid angle $d\Omega$ about the direction $\theta, \phi$ at the distance characterised by $z$; $F$ is the detection probability for these sources (known if we can estimate the selection effects in operation). The quantity $n(z, \theta, \phi)$ on the light cone leads to the estimate $\mu_{\text{lum}} = Mn$ for the amount of matter represented by these sources, where $M$ is the mass per source observed.

(v) However, as there may be considerable 'hidden matter', the observed matter may only make a small contribution to the total energy density present. It is therefore important that we can in principle also estimate the local energy density present at points on the past null cone through its local dynamic effects (e.g. by observing galaxy rotation curves or from virial theorem estimates), leading to estimates $\mu_{\mathrm{dyn}}$. Given the values of these observable quantities we have to estimate the total energy density $\mu(z, \theta, \phi) \equiv \mu_{\mathrm{total}}$ on the past light cone, which is a field on the light cone that determines the spacetime curvature through the Einstein equations. It is possible that there are present large amounts of spatially homogeneous matter and energy not detectable by either method.

(vi) We can in principle determine the background black body radiation temperature $T(z, \theta, \phi)$ not only at the vertex p ('here and now') but also at distant points on our past light cone, through its effect on matter present there (e.g. through molecular lines in high-redshift absorption systems). This is then a field defined on the light cone, not giving direct information about its geometry but determining significant physical conditions there.

(vii) We can hope to obtain information such as element abundances and stellar ages not only 'here and now' but also at events down our past light cone. Such quantities are also fields defined on the light cone, not giving direct information about its geometry but determining physical conditions there. From this information we can deduce conditions at events in the distant past near the world line of the matter observed (Hoyle 1960, Ellis 1975).

### 3.3. *The standard model in observational coordinates*

The standard model is a Robertson–Walker spacetime with perfect fluid matter source characterised by the equation of state $p = (\gamma - 1)\mu$.

*3.3.1. The spacetime.* In observational coordinates $(w, r, \theta, \phi)$ the metric is

$$\mathrm{d}s^2 = R^2(\tau)[-\mathrm{d}w^2 + 2\,\mathrm{d}w\,\mathrm{d}r + g^2(r)(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\phi^2)] \tag{7a}$$

where

$$g(r) = (\sinh r,\ r,\ \text{or}\ \sin r) \tag{7b}$$

according as $k = +1$, 0, or $-1$, and

$$\tau = w - r \tag{7c}$$

(this is clearly a special case of (1)). The matter 4-velocity is

$$u^\alpha = \frac{1}{R(\tau)}\,\delta^\alpha_0 \qquad (\Rightarrow u^r = u^\theta = u^\phi = 0) \tag{8}$$

(in this case, $r$, $\theta$ and $\phi$ are all comoving). The matter energy density $\mu$, number density $n$ and radiation temperature $T$ are

$$\mu = \mu_0(R_0^{3\gamma}/R^{3\gamma}(\tau)) \qquad n = n_0(R_0^3/R^3(\tau)) \qquad T = T_0(R_0/R(\tau)) \tag{9}$$

where $\mu_0$, $n_0$, $T_0$ are constant except during short time intervals when significant matter–radiation interaction takes place, mainly at very early times.

The conformal time coordinate $\tau$ used here is related to the usual time coordinate $t$ by

$$t = \int R(\tau)\, d\tau.$$

The scale function $R(\tau)$ has the standard Friedman form depending on the equation of state; for example at late times when the pressure is zero, $\gamma = 1$ and if $k = +1$

$$R(\tau) = Q_0(1 - \cos\tau) \qquad t(\tau) = Q_0(\tau - \sin\tau) \tag{10a}$$

where $Q_0$ is determined in terms of the deceleration parameter $q_0$ and Hubble constant $H_0$ by

$$Q_0 = q_0/H_0(2q_0 - 1)^{3/2} \tag{10b}$$

(with similar expressions if $k = 0, -1$). The present values of $\tau$ and $R$ (assuming $k = +1$) are

$$R_0 = \frac{1}{H_0(2q_0 - 1)^{1/2}} \qquad \tau_0 = \cos^{-1}\left(\frac{1 - q_0}{q_0}\right). \tag{11}$$

Similar expressions hold when $k = 0$ and $-1$.

*3.3.2. Observable quantities in the standard model.* Assuming $p = 0$, $\Lambda = 0$, the observable quantities on the null cone $C^-(p)$ (which is the null surface $w = w_0$) in a FLRW universe (7) are given by

$$1 + z = R_0/R(\tau) \tag{12a}$$

$$r_0 = R(\tau)g(r) \tag{12b}$$

and $N_0$, $\mu$, $T$ given by equations (6*a*), (9), with $r$ related to $\tau$ by

$$\tau = w_0 - r. \tag{12c}$$

These lead to the observable relations

$$r_0(z) = \{q_0 z + (1 - q_0)[1 - (1 + 2q_0 z)^{1/2}]\}/H_0 q_0^2(1 + z)^2 \tag{13}$$

$$\mu(z) = 6H_0^2 q_0(1 + z)^3 \qquad T(z) = T_0(1 + z) \tag{14}$$

$$N_0(z) = n_0 R_0(2q_0 - 1)^{1/2}(1 + 2q_0 z)^{-1/2}(1 + z)^3 \tag{15}$$

where $N_0(z)$ (in this case, isotropic) is defined by (6*b*), $H_0$ and $q_0$ define the FLRW geometry and the time of observation, and $n_0$, $T_0$ are constant during most of the evolution of the universe ($T_0$ will be constant at times when $p = 0$; however $n_0$ will vary during times of significant source evolution, e.g. when galaxies are forming).

The redshift $z$ is related to the coordinate $r$ on the null cone $C^-(p)$ by

$$1 + z = (2q_0 - 1)/q_0(1 - \cos(w_0 - r)) \tag{16}$$

if $k = +1$ (with similar expressions if $k = 0$ or $-1$).

## 4. Null data best-fitting

We discuss this in stages, as outlined in § 1.4 (cf figure 2).

### 4.1. *Choice of vertex points*

First we have to specify the choice of corresponding points p, p′ in U, U′ as vertices for the comparisons. Regarding U as representing the real universe, the point p in U is the point 'here and now' which is the single event from which it is possible for us to make cosmologically relevant astronomical observations (Hoyle 1960, Kristian and Sachs 1966, Ellis 1975, 1980). The point is that the whole history of astronomical observations (say 2000 years) is negligible relative to the scale of the visible universe (of the order of $10^{10}$ years); and if we were now to embark in a spaceship and travel for 2000 years in order to view the universe from another spatial vantage point, we would not even succeed in leaving our own galaxy; but a cluster of galaxies is a point in a cosmological model. Thus on a cosmological scale, we cannot choose the spacetime event from which we observe the universe: it is given to us.

It does not matter what point we choose as the corresponding point p′ in the FLRW universe model U′, because of the spatial homogeneity of that universe. Conditions are equivalent at all points, except for the variation with time. However we cannot *a priori* specify the appropriate (FLRW) time parameter $t_0$ at that event, for this effectively defines the time of observation, which we regard as one of the final products of comparison of theory and observations (effectively this will be the observationally determined age of the universe). Thus this quantity will be a free parameter, whose value can only be determined at the final stage of the fitting procedure. Apart from this, it does not matter what point in U′ we choose as p′. In summary: in the particular situation we are considering, the choice of comparative points p, p′ is not a problem: p is fixed in U and p′ can be chosen arbitrarily in U′ (because U′ is a FLRW universe). This feature saves us from having to compare all pairs of points p, p′ to find the appropriate identification.

### 4.2. *Fitting the 4-velocity*

The next step is that we have to fit the matter 4-velocities $u^a$, $u'^a$ representing the motions of 'fundamental observers' at the points p, p′ in the two models. Now in the FLRW universe U′, there is a unique† 4-velocity $u'^a$ determined by the spacetime geometry at each point (and so in particular at p′), namely the 4-velocity orthogonal to the surfaces of homogeneity. This is the only 4-velocity about which a FLRW universe appears to be isotropic. In the lumpy model U, the 4-velocity $u^a$ at p should be chosen to correspond as closely as possible to this situation, so it should isotropise observations about p as much as possible.

The general procedure would be to construct an overall measure $M(A)$ of the anisotropy $A$ of all cosmologically relevant astronomical observations on the past light cone of p, weighting each of them according to their reliability, and then choose $u^a$ at p to minimise this measure. Specific measures of anisotropy can be derived from galaxy number counts or mass estimates and used to derive related average 4-velocities (Ellis 1971).

However in practice the anisotropy of one of the observational quantities, namely the temperature of the microwave background radiation, is so much better determined than all the others that it seems appropriate to follow general practice and use this alone to determine the preferred 4-velocity. In this case one obtains a unique choice of $u^a$ at p by eliminating the microwave background radiation dipole component

† Provided $\mu + p \neq 0$.

(Sachs and Wolfe 1967). This choice is appropriate not only because of the superior experimental status of the microwave anisotropy, but also because this radiation comes from very large distances (a redshift of at least 10, and possibly 1000) and so gives an isotropy measure based on averaging the properties of the universe over very large scales. This will be better than more local averages (e.g. based on the velocity of observed matter within the local supercluster) which will be more strongly affected by local deviations from isotropy.

The proposal, then, is to determine the average 4-velocity $u^a$, $u'^a$ at the points p, p′ in U, U′ as that which maximises the isotropy of the microwave background radiation. In the FLRW model, this will give the usual choice; in the 'lumpy' model, a well determined 4-velocity. If the FLRW model gives a good fit, the remaining dipole anisotropy in all the other quantities should then be close to zero; for example, number counts should show no dipole isotropy about the fitted velocity $u^a$. This gives a consistency check on the velocity-fitting performed through the microwave background radiation. In effect, one splits the measure $M(A)$ into two parts: $M1(A)$, the measure of anisotropy in the microwave background radiation, and $M2(A)$, the anisotropy in all other quantities. The first is used to make a unique choice, and the second to verify this choice. Should there remain substantial dipole anisotropies in cosmological data as measured by $M2(A)$ once the choice dictated by $M1(A)$ has been made, then either the microwave background radiation is not cosmological and should not be used to determine the rest frame at p, or a FLRW model cannot give a good fit. This test of goodness of this choice will in effect be automatically made through the procedures discussed in the next section: a poor choice of 4-velocity will be apparent as a dipole term in the goodness-of-fit criteria, and this feature can be made explicit when examining those criteria. Nevertheless it is interesting to isolate the kinds of effect expected due to 'wrong' choice of 4-velocity. We can do so by considering observational relations in a FLRW universe model U′ observed from a 4-velocity $v^a$ different from the unique 4-velocity $u'^a$ defined by that model.

The basic result of such a 'wrong' choice of 4-velocity is threefold. Firstly, there is a dipole effect on observed redshifts, resulting in a change of all observed bolometric intensities by a factor of $(1+z_v)^4$ and specific intensities by a factor $I(\nu(1+z_v))/(1+z_v)^3$ where $I(\nu)$ is the spectrum observed and $z_v$ the redshift caused by the 'wrong' choice of 4-velocity at p′. Thus intensities change effectively by a $\cos^3\theta$ law rather than $\cos\theta$. Secondly, observed galactic 4-velocities are changed by a $\cos\theta$ transformation over the 2-sphere. Thirdly, image shapes and sizes are left unchanged, so there is no resultant distortion due to bad choice of 4-velocity, but solid angles are changed by a factor $(1+z_v)^2$ because the length scale determining the unit sphere (on which solid angles are defined) is changed by the length-contraction effect. Consequently area distances also change according to this rule, i.e. by a factor $(1+z_v)^2$. The first and third effects imply that (magnitude, redshift) curves and number counts will show an anisotropy of a somewhat complex nature.

Apart from helping understand the situation when a 4-velocity fitting has been badly made in an almost-FLRW universe, this analysis is of interest because it describes the situation when a FLRW universe is viewed from a galaxy measured to be moving relative to the microwave background radiation (which is the case for our galaxy). Therefore if the real universe is close to a FLRW model, we do indeed expect to determine such anisotropies in observations made from the earth. Thus for example if the standard interpretations are correct we should measure an anisotropy in radio source counts (Ellis and Baldwin 1984).

### 4.3. *Choice of comparison points on the null cones*

Having chosen the vertex points p, p′, the null cones $C^-(p)$, $C^-(p')$ are uniquely determined; the intention is to compare initial data on them. To do so we make a choice of correspondence between points q, q′ on the two null cones. We can then determine the difference between data at each such pair of points, and integrate this difference (suitably weighted) over the observable region of the null cone by letting q range over $C^-(p)$ down to the relevant limiting distance $y^*$ in each direction (then q′ ranges over the corresponding region of $C^-(p')$).

*4.3.1. Choice of corresponding 2-spheres.* The choice of corresponding points q, q′ primarily involves choosing a measure of distance down each null cone; then equivalent points q, q′ will lie on the 2-spheres S, S′ at the same distance down each light cone from its vertex. It is here that a significant choice has to be made: there are different possible measures of distance down the null cone, and the 2-spheres obtained will depend on this choice. The geometrically preferred distance is an affine parameter $v$ for the null geodesic vector $\boldsymbol{k}$ generating the null cone (i.e. $k^a = \mathrm{d}x^a/\mathrm{d}v$) normalised at the vertex by the condition: $\boldsymbol{k}\cdot\boldsymbol{u}|_\mathrm{p} = -1$. However this cannot be directly determined by astronomical observations from p, so it is not an observable measure of distance. The next preferred choice is the area distance $r_0$ representing the spreading out of the light cone from the vertex, which is in principle directly measurable (Ellis 1971). Unfortunately, however, in practice it is very difficult to measure this quantity reliably. The third best option is the observed redshift $z$, which has the great advantage of being (relatively) easily measurable (it can be determined directly from observed source spectra, provided an identifiable spectral line is present). The disadvantage is that it does not directly represent distance down the null cone; rather the observed value $z$ is related to the cosmological redshift $z_\mathrm{c}$ by the relation

$$1+z=(1+z_\mathrm{O})(1+z_\mathrm{c})(1+z_\mathrm{S}) \tag{17}$$

where $z_\mathrm{O}$ is the redshift due to the peculiar velocity of the observer O (at p) and $z_\mathrm{S}$ that due to the peculiar velocity of the source S (at q). The latter cannot be directly distinguished from the cosmological redshift by any local observations at p. Thus the presence of random velocities (relative to the motion of a 'fundamental observer') means that use of redshift as a distance measure will be misleading: galaxies with the same observed redshift may be at different distances from the vertex.

Each choice has advantages and disadvantages. However in the context of an observationally based fitting procedure, it seems inevitable we have to use the cosmological redshift $z_\mathrm{c}$ to estimate distance down the null cone, because the observable redshift $z$ is so much easier to measure than the other quantities; and equation (17) provides a basis for estimating $z_\mathrm{c}$ from $z$. Nevertheless given this choice, it must continually be reassessed, particularly because there is in the literature controversy over use of $z$ as a distance measure (e.g. in the writings of Arp and Burbidge) and we are essentially accepting the standard view. It is also at this point that the importance of choosing the correct 4-velocities (in the previous step) becomes clear; for if we change that choice, the 2-spheres S, S′ we determine will also alter, whichever distance measure we use. In the case of redshifts this is clear from equation (17): changing the choice of 4-velocity at p will alter $z_\mathrm{O}$ in a dipole manner at q, and so alter the 2-sphere S given by constant values of $z_\mathrm{c}$ determined from measures of $z$ and estimates of $z_\mathrm{S}$.

*4.3.2. Choice of corresponding points in each 2-sphere.* Having determined corresponding 2-spheres S, S′, it remains to determine which points q, q′ correspond in S, S′. Here the isotropy of the FLRW universe simplifies matters; for it implies that all points on the FLRW 2-sphere S′ are equivalent. Thus given any point q in S, it does not matter which point q′ in S′ we choose to correspond with it, for conditions are the same at all of them. This saves us having to consider all possible pairs of points q, q′ in our fitting process.

What this means in practice is that once we have chosen observational coordinates $(z_c, \theta, \phi)$ on the null-cone $C^-(p)$: $w = w_0$ in S and corresponding coordinates in S′, we can simply compare observational quantities at the same coordinate values in S, S′, and this corresponds to comparing them with the desired point identification. Thus we can simply integrate a suitable measure of the difference between initial data in the two spaces, as measured in observational coordinates, over the visible region of the null cone.

### 4.4. Fitting the null data

Observational quantities in the lumpy universe U are characterised by equations (2)-(6), and in the FLRW model U′ by equations (9)-(16); and it is known (Ellis *et al* 1985) that these quantities form the null data for cosmology. The desired best fit should minimise the integrated difference between these data in the two spacetimes in some suitable measure, where the measure weights the data with reliability and hence with distance from the point of observation (the vertex of the null cone).

However, there is an important distinction to be made. Some of the data (the transverse velocity components and distortion) are identically zero for all FLRW universes. Thus the FLRW models cannot be adjusted to give a best fit to the lumpy model in regard to these quantities (there is no corresponding FLRW parameter to adjust). Similarly one cannot adjust the FLRW model to give density variations on the 2-spheres S′. Hence the fitting procedure now separates into two steps: considering in turn isotropy and homogeneity. In effect this means we first isotropise on the past null cones $C^-(p)$, $C^-(p')$ to obtain averaged (spherically symmetric) geometries; and then adjust the (spatially homogeneous) FLRW universe U′ to give a best fit to the spherically-symmetric average of the null data in the lumpy universe.

*4.4.1. Anisotropic null data.* In a FLRW universe, all quantities are isotropic about every observer. Given a measurable quantity $f$, in order to proceed with a detailed fitting we need to be able to combine it with some other observable quantity to obtain an observable relation; usually the second quantity may be taken as the redshift $z$, so the observable relation is $f = f(z)$. The basic procedure for dealing with anisotropies is to integrate each observable relation $f(z, \theta, \phi)$ over each 2-sphere S (characterised by a redshift $z$), to determine a (spherically symmetric) average

$$\bar{f}(z) = \int_S f(z, \theta, \phi)\, \mathrm{d}\Omega \tag{18a}$$

and an anisotropy measure

$$\hat{f}(z) = \int_S (f - \bar{f})^2\, \mathrm{d}\Omega \tag{18b}$$

where $d\Omega = \sin\theta\, d\theta\, d\phi$. Then

$$\Xi(f) = \int_0^{z^*} W_f(z)\hat{f}(z)\, dz \tag{18c}$$

is a number measuring the overall lack of isotropy in the quantity $f$, where $W_f(z)$ is a weighting function proportional to the quality and reliability of the data for $f$ (allowing suitably for decrease of quality of data with increasing redshift, cf Ellis 1980). In the FLRW case, $\hat{f}(z) = 0$ and $\Xi(f) = 0$, so we should be able to choose limits $\lambda_f(z)$, $\Lambda_f$ such that if either

$$\hat{f}(z) > \lambda_f(z) \tag{18d}$$

for some $z$, or

$$\Xi(f) > \Lambda_f \tag{18e}$$

we will reject *all* FLRW models as a reasonable representation of the data for $f$.

As mentioned above, there are two rather different applications of this test. Firstly, the proper motions $M^A$ and the distortion $f_{AB}$ (see equations (3) and (5)) are identically zero in a FLRW model. Applying the above prescription, in these cases $\bar{M}^A$ and $\bar{f}_{AB}$ are both identically zero, so

$$\hat{M}^A(z) = \int_S (M^A)^2\, d\Omega \tag{19a}$$

$$\hat{f}_{AB}(z) = \int_S (f_{AB})^2\, d\Omega \tag{19b}$$

are direct measures of the non-FLRW nature of the universe. Secondly, the above test can be applied to all the measurable non-zero relations or quantities, i.e. area distance estimates $r_0(z)$, number counts $N_0(z)$, density estimates $\mu(z)$ and the background radiation temperature $T(z)$ (in particular applying it to the latter at $z = 0$, i.e. to measurements made at the present time). One could of course apply a harmonic analysis to the anisotropies detected, instead of the simple anisotropy measures (18*b*, *c*). In particular if there were a substantial consistent dipole effect, this would indicate that the 4-velocity at p had been chosen wrongly (cf § 4.2), and one could then try varying that choice to see if one could in this way decrease the anisotropy measures discussed here. The prime purpose of this harmonic analysis would be for comparison with other models in which these components might be non-zero (e.g. Bianchi universes).

In practice, in most cases we will only have data at a finite number of points on the light cone (corresponding to observations of discrete sources). Then we would in the obvious way replace the integrals above by sums over data at all the actual observational points. In this case, an appropriate choice for the weighting function $W_f(z)$ might be the inverse of the variance of the data.

In some cases, we might be able to measure a quantity $f$ but be unable to determine an observational relation because we could not determine other properties of the relevant sources (e.g. $f$ might be number counts of sources whose redshift could not be measured). In this case we clearly cannot examine the relation in depth as above, but can still look for overall anisotropies in the resulting observations of the (single) quantity $f$; any such anisotropies again represent a lack of FLRW character of these sources.

*4.4.2. Isotropic null data.* The spherically averaged relations $\bar{f}(z)$ (equation (18$a$)) define a spherically symmetric geometry. Examining this in depth down the light cones, we aim to find a 'best-fit' FLRW universe and a best-fitting of that universe. This provides the basis for tests of the degree to which the lumpy universe U is approximated by the FLRW model U′ on the past light cone, thus serving as a test for spatial homogeneity of the observed region of the universe.

Equations (13)-(15) give the observational relations expected in a FLRW model (with pressure and cosmological constant both zero). We want to fit these simultaneously to the data for $\bar{r}_0(z)$, $\bar{\mu}(z)$, $\bar{T}(z)$ and $\bar{N}_0(z)$ to determine the 'best-fit' FLRW model characterised by constants ($H_0$, $q_0$, $n_0$, $T_0$). In doing this, we want to weight the data with respect to precision, so that the more precise the data the more they will contribute to the fit; that is, we weight them by the appropriate functions $W_f(z)$ (see (18$c$) above). The method proposed is use of least squares weighted by the variance $\sigma_f$.

More precisely, the data sets available will be a series of spherically averaged observed values $\bar{r}_{0_i}$, $\bar{\mu}_i$, $\bar{T}_i$ and $\bar{M}_{0_i}$ of $r_0$, $\mu$, $T$ and $M_0$ at redshift values $z_i$ and with variances $\sigma^2_{r_i}$, $\sigma^2_{\mu_i}$, $\sigma^2_{T_i}$ and $\sigma^2_{M_i}$ respectively at these redshifts. We construct the $\chi^2$ function as usual:

$$\chi^2 = \chi_1^2 + \chi_2^2 + \chi_3^2 \tag{20$a$}$$

where

$$\chi_1^2 = \sum [\sigma_{r_i}^{-2}(\bar{r}_{0_i} - r_0(z_i))^2 + \sigma_{\mu_i}^{-2}(\bar{\mu}_i - \mu(z_i))^2] \tag{20$b$}$$

$$\chi_2^2 = \sum [\sigma_{T_i}^{-2}(\bar{T}_i - T(z_i))^2] \tag{20$c$}$$

$$\chi_3^2 = \sum [\sigma_{M_i}^{-2}(\bar{M}_{0_i} - M_0(z_i))^2]. \tag{20$d$}$$

Here $r_0(z_i)$, $\mu(z_i)$, $T(z_i)$ and $N_0(z_i)$ are values *calculated* from equations (13)-(15). Now in principle we find the optimum values of $H_0$, $q_0$, $n_0$ and $T_0$ by minimising $\chi^2$ with respect to each of these parameters simultaneously:

$$\frac{\partial}{\partial q_0}\chi^2 = 0 \qquad \frac{\partial}{\partial H_0}\chi^2 = 0 \qquad \frac{\partial}{\partial n_0}\chi^2 = 0 \qquad \frac{\partial}{\partial T_0}\chi^2 = 0 \tag{21}$$

yielding values of the parameters $H_0$, $q_0$, $n_0$, $T_0$ characterising the best-fit FLRW model (that is, FLRW spacetime and matter parameters) for the data set. The resulting minimum value for $\chi^2$ gives a test for goodness of fit. It is worth remarking one could carry out this test directly on the original data $f$ rather than their spherical average $\bar{f}$; in the case where the data are nearly spherically symmetric this should make very little difference.

In practice, it may turn out that it is better to treat $H_0$ as a scaling factor determined from a complex series of local observations (cf Rowan-Robinson 1985) instead of a parameter to be determined from these equations. Furthermore, in this approach, we have assumed the parameters are independent, but there may be relations between them. If relations between the parameters are known, they can be used to eliminate some parameters in favour of others; but in general such relations will involve additional constants that have to be determined observationally. As an example, we need to consider carefully the relation between the number-count constant $n_0$ and the density constant $\mu_0$. If there is a lot of 'dark matter' present these may effectively be independent; on the other hand they may be related by a total-mass per source parameter $M_s$ so that $\mu = M_s n$. If so, this relation should be added to those above, so the parameters are constrained by the relation

$$\mu_0 = 6H_0^2 q_0 = M_s n_0 \tag{22}$$

where $M_s$ has to be determined by a combination of theory and extra observations. In a similar way the temperature $T_0$ may be related in some theories to $M_0$ by a constant (essentially the entropy per unit mass), and this constant may be constrained through further theory and related possible observations, e.g., the theory of nucleosynthesis and observations of element abundances. In such cases, this theory can be used in conjunction with appropriate observations to give further contributions to the value of $\chi^2$.

The procedure outlined here is nothing other than a version of the way astronomers already determine FLRW parameters on the past light cone, which is as it should be. In reality the procedure will be very complex because of astrophysical complications (see, e.g., Sandage and Tammann (1986) for a recent survey). The most important problem is that of determining the intrinsic luminosity or size of very distant sources (far down the light cone), because that is the basis from which observations can be used to determine the area distance; and this estimation is bedevilled by the problem of an unknown source evolution. It is therefore very important that in recent years the use of supernovae as distance indicators is becoming feasible (Wagoner 1977, van den Bergh and Pritchett 1986), because the intrinsic brightness of a supernova can be estimated from physical theory and so avoids the evolution issue. The particular significance is that this saves us from having to use a circular argument whereby the intrinsic source brightness can only be estimated by comparison with nearby sources plus estimates of the source evolution, which estimates can only be carried out on the basis of an already fitted FLRW model.

*4.4.3. Fitting the spacetimes.* Having obtained the best-fit set of parameters by suitable observational procedures on the past null cone, as outlined above, we are effectively in the position of the geodicist who has determined a value for the radius of the Earth. He still has to specify the best-fitting of a sphere of that radius to the 'lumpy' Earth; we still have to determine the best-fitting of the FLRW model characterised by the parameters just determined, to the 'lumpy' real universe.

The FLRW model is related to observations in the following way. Equation (7) gives the spacetime metric on the past null cone in observational coordinates, where $R(\tau)$ is determined from $(12a)$:

$$R(\tau)=R_0/(1+z) \tag{23}$$

and (if $k=+1$) $r$ is determined by (16):

$$r=\tau_0-\cos^{-1}\left(1-\frac{q_0(1+z)}{2q_0-1}\right) \tag{24}$$

where $R_0$ and $\tau_0$ are given by (11) (note that $\tau_0=w_0$ because $r=0$ at the point of observation). Similar relations hold if $k=0$ or $-1$. The matter 4-velocity (8) is then determined, as is the energy density $\mu$, number density $n$ and radiation temperature $T$ (see equation (9)), the constant $\mu_0$ being determined from (14):

$$\mu_0=6H_0^2q_0. \tag{25}$$

These relations together completely specify the spacetime and the matter on the past null cone, in terms of the observational coordinates.

The lumpy universe model, on the other hand, is given by the spacetime metric (1) and matter description determined by $u^a$ and $n$. The metric components $g_{AB}$ are observationally measurable (cf (4) and (5)) but the components $g_{0a}$ are not; however, given the other observational quantities (determined by (2)-(6)) these metric components can be determined by integrating the field equations down the light cone from the vertex (Ellis *et al* 1985). Thus in principle we can determine the lumpy universe model on the past null cone direct from the data (in practice one will need to discretise the representation and interpolate data, but essentially that is a detail rather than a matter of fundamental principle).

Consequently we can determine the universe models U, U′ on the corresponding past null cones $C^-(p)$, $C^-(p')$ from the observational data by the procedures outlined above. They will both be given in observational coordinates $(z, \theta, \phi)$ on that null cone; and corresponding points in the two universes (as discussed in §§ 4.1-4.3) will be given by the same coordinate values. Thus not only has the FLRW model been determined but a fitting has been specified between U and U′ (determined by the observational coordinates), and their properties at corresponding points can be compared directly.

### 4.5. *Fitting off the null cone*

The fitting has now taken place on the null cone $C^-(p)$, on the basis of the observable data for cosmology on $C^-(p)$. Finally we need to construct the best-fit spacetime off the null cone (to the future and past) from the data on this surface and, insofar as this is possible, try to estimate how good the fit is off the light cone.

Because the FLRW universe U′ is rigidly determined by its initial data (provided the equation of state is known), the best-fit of FLRW data analysed in the previous section determines the FLRW universe off the past null cone as well as on it. The metric is given by (7) with $R(\tau)$ determined by (10), the matter 4-velocity by (8), and the energy density and background radiation temperature by (9) and (14). The null cone itself is given by $w = w_0$; because of the choice of coordinates (comoving with the fluid flow, with $w$ a conformal time coordinate) the extension off the null cone simply consists in letting $(w, z)$ vary over a suitable range of values around $(w_0, 0)$, the equations remaining unaltered. This 'rigid' evolution prediction for the FLRW model (of course corresponding to the usual way a FLRW model evolves) is equally valid to the future and the past of $C^-(p)$ (figure 4($a$)).

In the case of the lumpy universe model U, the field equations can again be integrated off the null cone on the basis of the observable data on it (Ellis *et al* 1985), but in general there are no simple analytic formulae for the resulting spacetime. In the nearly-FLRW case the perturbed field equations may be amenable to analytic investigation, but in general numerical analysis will be needed. Thus given the data on $C^-(p)$ determined by astronomical observations, suitable numerical programs should allow comparison of the spacetimes U, U′ off the null cone as well as on it.

There are in fact two rather different regions involved. Inside the past null cone (i.e. for $w < w_0$) the major problem is simply the issue of numerical integration, made difficult by the data being on a cone (with coordinate singularities necessarily arising at the vertex). The solution in this region is uniquely determined from the data on the null cone. However there are problems *of principle* in integrating to the future of the past light cone $C^-(p)$ $(w > w_0)$ for we do not have sufficient data to do so. Such events lie outside the domain of dependence of $C^-(p)$ (see figure 4($b$)), and conditions there are not uniquely determined from data on the null cone; for example, gravitational

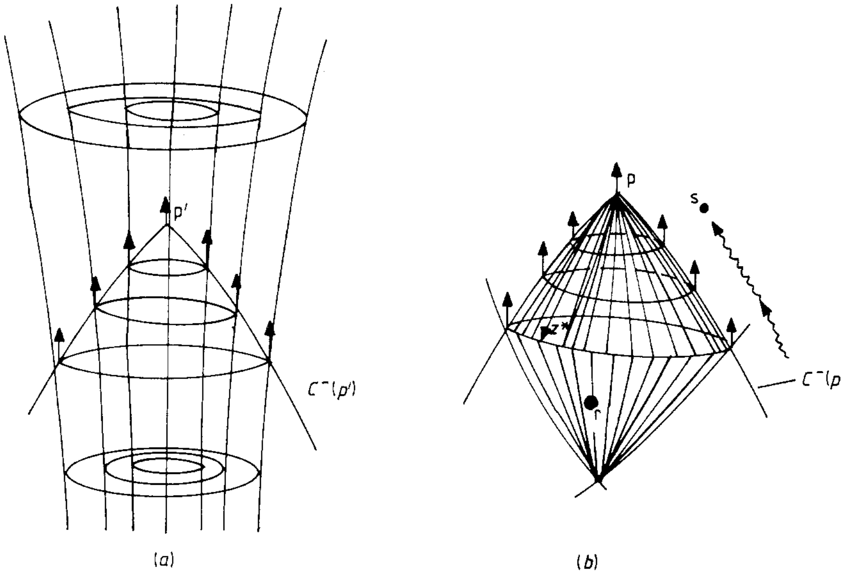


**Figure 4.** (*a*) In a FLRW universe, initial data on a section of the past light cone of p′ rigidly determines the evolution of the universe to the future and past of that light cone. (*b*) In a lumpy universe, data known to some limiting redshift $z^*$ on the past light cone of p determines the spacetime uniquely in the past domain of dependence of the region of the null cone limited by $z^*$ (e.g. at the point r), but cannot uniquely determine the spacetime at points outside this domain of dependence, e.g. at the point s to the future of the past null cone (in principle a gravitational wave from a source with which we have not yet had any causal contact can destroy any prediction we make about events at s).

waves from as yet unseen objects could conceivably invalidate any predictions we care to make (Ellis 1984). Thus† we can only estimate conditions to the future of the null cone by making some kind of 'no-interference' assumption, that the partial data we do have for conditions in this region will not be invalidated by interference from events and objects about which we have no evidence whatever. This kind of assumption underlies even predictions such as that the Sun will rise tomorrow and that there will be an eclipse of the moon next year.

Given some such assumption, we can in principle estimate conditions both to the future and the past of the null cone and compare the predictions of the best-fit smoothed-out FLRW model U′ for these spacetime regions with those of the more realistic lumpy universe model U. The case when U can be thought of as a perturbed FLRW model is being examined at present; such investigations are needed to extract the full implications of the best-fit procedure suggested in this paper.

### 4.6. *Averaging on the null cone*

The procedure proposed above fulfils most of the criteria one would want for a good fitting procedure and incorporates aspects of many of the alternatives (see the appendix). However there is one significant lack: it does not explicitly include the idea of

† Unless we live in a 'small universe' that we have already seen around, see e.g., Ellis and Schreiber (1986); in that case we do have sufficient data to predict to the future in principle.

averaging, discussed in § 2 above. Can we somehow relate that aspect of 'best-fitting' to the presently considered null data approach?

We can do so in essence by dividing up the past null cone $C^-(p)$ into sampling volumes $V_{\Delta x,q}$ centred on points q (with coordinates $w_0$, $z$, $\theta$, $\phi$) of $C^-(p)$ and characterised by the coordinate increments $\Delta x = \{\Delta z, \Delta\theta, \Delta\phi\}$ about q (figure 5). Then, as in § 2.2, we can associate a FLRW model $U'(q, \Delta x)$ with the comoving time development of this volume and, as in that section, we can consider the family of FLRW models obtained as q varies in $C^-(p)$ and so see how consistently the same FLRW universe is indicated by averaging over different sampling volumes.

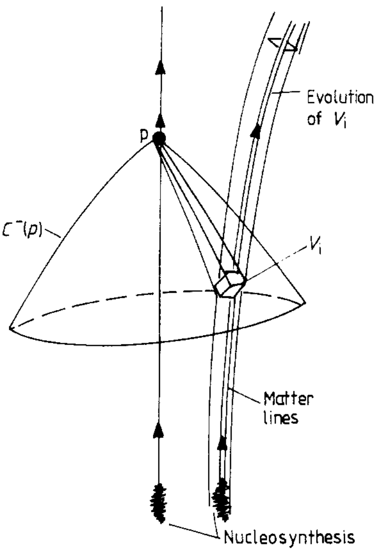


**Figure 5.** Observations of data in an averaging volume on the past light cone can be used to determine conditions in the past and future of that volume if we assume it evolves like a FLRW universe. In particular, element abundances determine conditions in the very distant past near the world lines of the matter observed.

In practice, what this amounts to is a variant of the averaging procedure for the data suggested in § 4.4 above. Actual observations of sources will result in discrete data (a set of figures for each source), which for purposes of analysis will be kept in a suitably organised catalogue. The essential step is grouping the data, i.e. identifying each source as belonging to one or other of a chosen set of averaging volumes $V_i$ covering the observable region of the null cone, where $i$ is an index for these volumes. As suggested above, it might be suitable to use the coordinates $(z, \theta, \phi)$ as the basis for defining them, but this must be continually reviewed, for if inappropriate the resulting model is wrong; in essence this is the basis of the dispute between Arp and other astronomers (is the observed redshift $z$ really a suitable coordinate to use for defining averaging volumes?). Thus the basis for the method is essentially the identification of clusters of objects.

In detail, a catalogue would be set up containing the relevant data for cosmology that can be obtained from observation of distant objects (seen in emission or absorption). Thus it would contain: ($a$) an identifier for the object; ($b$) angular coordinates $\theta$, $\phi$ for its position; ($c$) the observed redshift $z$; ($d$) an area distance estimate obtained from angular sizes or apparent luminosity measures; ($e$) an image distortion measure; ($f$) a source proper motion measure†; ($g$) a number count measure; ($h$) estimates of the matter present from dynamical or absorption measures; ($i$) measures of the black body radiation temperature at the object; ($j$) measures of element abundances at the object; ($k$) other relevant information where attainable, e.g. estimates of ages of the observed objects. In general for any particular source one would only be able to obtain some of these data, but the aim would be to obtain as complete as possible information for each object, and to do that for objects that gave a good coverage of the whole sky.

Having grouped the observed objects into a set of averaging volumes $V_i$ on the basis of the observational data (principally the apparent position in the sky and the redshift, but modified by other distance indicators if necessary) one can then fit a FLRW universe to each volume by effectively the method outlined above (use of the $\chi^2$ test on the observations in each volume, but using additionally data such as the element abundances and ages to constrain the parameters where practicable, cf e.g. Gott *et al* 1974, Sandage and Tammann 1986). A table of FLRW parameters $U_i' = \{q_{0i}, H_{0i}, \mu_{0i}, n_{0i}, T_{0i}\}$ will be obtained as a result, plus estimates of the closeness to FLRW geometry in each volume (determined from anisotropy measures within that volume). One can then survey the whole set of FLRW data for consistency, and also average it over 2-spheres at the same depth so as to effectively do the analysis suggested above (first isotropising and then finding a single best-fit FLRW model in depth, using a suitable weighting of the data with distance and reliability). What is happening here is we utilise the fact that FLRW predictions on the null cone are rigid: they are determined by just two or three numbers (as emphasised by Sandage). Thus the separate FLRW models giving a best fit at different parts of the null cone must all agree in order to give a single best-fit FLRW model.

It is important to notice two features of this fitting. Firstly, assigning say a value of $q_0$ to a volume $V_i$ from an area distance measurement is an essentially non-local procedure: the assignment is made on the basis of the non-local relation between $V_i$ and the vertex q (the area distance $r_0$ measures the total spreading of the light-cone between q and $V_i$, depending on the entire spacetime between them). The other measurements determine local properties of the universe at the observed region $V_i$ (identified by the redshift $z$) and parameters assigned on the basis of those measurements are descriptive of local conditions at that region. Secondly, measurements of features such as element abundances do not directly determine any geometrical property of the past light cone; however, they do determine conditions in the very distant past near the comoving volume represented by $V_i$ (figure 5). The simplest way to investigate the implications of these data is to attach a FLRW universe to this volume, using the element abundances to estimate its parameters; and this is precisely what is done in the 'averaging' procedure suggested here. One of the most important such features is the conditions that can be inferred from the microwave background radiation; in this approach, the essential element is that we can in principle determine those volumes on the null cone where decoupling takes place, if we can estimate the decoupling temperature $T_d$ and assume that formula (14) holds for $T$ all the way back to decoupling. However unfortunately life is not so simple: physical conditions at decoupling are quite uncertain, and decoupling could take place anywhere between a redshift of 1000

† We realise this is extremely difficult to measure. Nevertheless the fact remains that proper motions are essential data for observational cosmology because they are part of the initial data for a cosmological model which have to be specified on a null cone, and so they must be determined to the best of our ability.

and 10 (Stebbins and Silk 1986). The way to extract the full information available from this radiation in this approach has still to be determined.

As well as the best-fit FLRW parameters, we would hope to specify the scales on which this model gives a good description. This depends on estimating the physical size of the averaging volumes $V_i$; the problem arises that except in the case of nearby volumes, we can only estimate this size once the model is at least partially determined (the cross sectional area is determined from $\Delta\theta$, $\Delta\phi$ once the area distance is known, which is in principle measurable, but the distance corresponding to $\Delta z$ can only be estimated once the metric coefficient $\beta$ is known, and this has to be found by solving the field equations or estimated by fitting a FLRW model). Thus it seems inevitable that a recursive procedure is needed in order to divide the light cone into volumes $V_i$ with approximately equal physical sizes. Ultimately this is what we wish to do; for a single FLRW description will be good on some scales but not on others, and so for some choices of $V_i$ but not others. The analysis will determine this scale once we have divided the light cone up into appropriately sized volumes.

### 4.7. *The result*

What will the result be of the procedures suggested? Firstly, as with present approaches we will obtain a best-fit FLRW universe model utilising available data to the maximum. Secondly, we will obtain in addition a best fitting of that model to the more realistic 'lumpy' universe, determining corresponding spacetime points and so enabling comparison of geometric and physical features at corresponding points. Thirdly, we will obtain a statement of goodness-of-fit, which estimates the lack of FLRW nature of the lumpy universe as indicated by the available observations. This can in turn be compared with 'adequacy of fit' criteria to give an overall statement if it is reasonable to use a FLRW model to represent the lumpy model (at the chosen averaging scale) or not.

How could we choose appropriate criteria to use in such a test? On the one hand, we can compare the observed goodness of fit with that expected in various possible situations. Thus we would compare the fit obtained ($a$) with that given by other FLRW models; ($b$) with that given by FLRW models perturbed only by statistically expected fluctuations; ($c$) with that given by other models, e.g. the Bianchi spatially homogeneous models and the Tolman inhomogeneous models. On the other hand we can investigate the relation of goodness-of-fit to dynamics; in particular, one could try to estimate the maximum deviation from a FLRW model that would still allow the linearly perturbed Einstein field equations to give a good approximation to the non-linear dynamics of the full field equations in the 'lumpy' universe (cf D'Eath 1976).

As has been stated above, while this fitting procedure seems very natural once one takes the 'null fitting' viewpoint, its implications are not immediately clear. This aspect needs development; for example the implications can be investigated by examining how the fitting works in particular cases. Work is under way on this. The other point that needs mentioning is that the point identification proposed is via observational coordinates on the two past null cones, which is the natural choice to make in an observationally based approach. However problems may arise if there are caustics in the light cone of the lumpy universe model, for then the same world line may intersect the past null cone several times, and caustics will certainly be expected to occur on some averaging scales (the observed gravitational lensing of QSO images being evidence for their existence). Investigation is needed as to whether caustics would be significant on averaging scales suitable for a good fitting of U′ to U, and of how to handle them.

We have commented that the determination of parameters for a best-fit FLRW model proposed here is like the standard procedures already in use. A significant difference is our recommendation that further data (particularly, distortion measures and proper motions) be collected on a systematic basis, despite the obvious difficulties in so doing. The point is that whether we like it or not, these data are essential in determining how good a 'best-fitting' is. We should be clear as to how fitting procedures presently being carried out, such as using motions of the Virgo cluster (see, e.g., Sandage and Tammann 1986) and IRAS number counts (Yahil *et al* 1986) to determine the local density and so the deceleration parameter, relate to our proposal. These are essentially local density measurements determining $\mu$ at the light cone vertex p. They are of course very valuable; the present analysis highlights that (on a cosmological scale) the measurements are made essentially at a single point on the light cone. Their significance needs confirmation by data at other points, if it is attainable.

## 5. Local physics in an almost-FLRW universe

Normal coordinates can be used as the basis of an approach to the fitting problem (see the appendix), but here we are concerned with their use in investigating local physics in an almost-FLRW universe. These issues are intimately related, as a fitting procedure underlies any investigation of local physics in a curved spacetime (cf § 1.5).

One of the simplest fitting procedures is to use one of the family of 'normal coordinates', which provide a natural mapping between the curved spacetime and flat spacetime. The metric components of an arbitrary spacetime in locally geodesic coordinates about a point p (Ehlers 1973, p 44) directly indicate the deviation from flatness of the spacetime. In detail, in such coordinates

$$g_{ab}(x^e) = g_{ab}(0) + \tfrac{1}{3}R_{acdb}x^c x^d + \mathrm{O}((x^a x^a)^{3/2}) \tag{26}$$

where p is given by $x^a = 0$. One can examine the effect of this curvature on local physics by evaluating its effect on free particle motion (the implied effect on more complex systems then following straightforwardly, see e.g. Misner *et al* 1972, box 37.1). Free particle motion near p is easily determined in an inertial quasi-Galilean coordinate system based on Fermi-propagated axes along a geodesic world line c (Ehlers 1973, p 48), and locally geodesic coordinates are a special case of such coordinates provided the metric $g_{ab}(0)$ at p is chosen to be the canonical flat space metric $\eta_{ab}$ (corresponding to the use of an orthonormal basis there). The equation in these coordinates for free particle motion near p is

$$\mathrm{d}^2 x^\alpha/\mathrm{d}\tau^2 \approx -R^\alpha{}_{0\beta 0}x^\beta \tag{27}$$

(see Ehlers (1973, equation 2.(50)) or Misner *et al* (1972, equation (37.3)); of course this is really just a form of the geodesic deviation equation).

Now we can characterise a FLRW universe model by the following result.

*Lemma.* A cosmological model $\{M, g_{ab}, u^a, \mu, p, \Lambda\colon u^a u_a = -1, \mu > 0, p = p(\mu), \Lambda = \text{constant}\}$ is a FLRW universe if and only if

$$\begin{aligned} R_{abcd} = {} & \tfrac{1}{2}\kappa(\mu + p)(u_a u_c g_{bd} + u_b u_d g_{ac} - u_a u_d g_{bc} - u_b u_c g_{ad}) \\ & + \tfrac{1}{3}(\kappa\mu + \Lambda)(g_{ac}g_{bd} - g_{ad}g_{bc}). \end{aligned} \tag{28}$$

*Proof.* This follows because (28) is algebraically equivalent to the conditions (i) the Weyl tensor vanishes and (ii) the matter stress-energy tensor, determined from the Ricci tensor through the Einstein field equations, takes the form of a perfect fluid with 4-velocity $u^a$, energy density $\mu$ and pressure $p$. However because we are assuming the existence of an equation of state $p = p(\mu)$, the Bianchi identities show that (i) and (ii) hold if and only if the universe model considered is a FLRW universe (see, e.g., Ellis 1971, p 136).

Equation (28) shows that

$$R_{abcd}u^b u^d = \tfrac{1}{3}h_{ac}[\tfrac{1}{2}(\kappa\mu + 3p) - \Lambda] = h_{ac}(\ddot{R}/R) \tag{29}$$

where $h_{ac} = g_{ac} + u_a u_c$ is the projection tensor perpendicular to the vector $u^a$ and the last equality follows from Raychaudhuri's equation for the FLRW universe (Ellis 1971, p 136). Substituting (29) into (27) with $u^a = \delta^a_0$, i.e. with the geodesic world-line c chosen as one of the fundamental world lines in the FLRW universe model, we find

$$\mathrm{d}^2x^\alpha/\mathrm{d}\tau^2 \approx x^\alpha(\ddot{R}/R). \tag{30}$$

The solution is

$$x^\alpha(t) = \left(C^\alpha + K^\alpha \int_{t_0}^{t} \frac{\mathrm{d}t}{R^2(t)}\right) R(t) \tag{31$a$}$$

where the constants $C^\alpha$, $K$ are given by

$$C^\alpha = x^\alpha|_0/R(0) \qquad K^\alpha = \dot{x}^\alpha|_0 - (\dot{R}(0)/R(0))x^\alpha|_0. \tag{31$b$}$$

This is the form of free motion of a particle in quasi-Galilean coordinates in a FLRW universe (relative to the geodesic c; it is of course just the general solution of the geodesic deviation equation in this case). The term proportional to $C^\alpha$ is the motion of comoving particles (at rest relative to the fundamental expansion); the other term represents motion of free particles moving relative to fundamental observers.

If we assume one of the usual functional forms for $R(t)$ we obtain the following results from (31$a$):

| $R(t)$ | $t$ | $t^{2/3}$ | $\mathrm{e}^{Ht}$ |
|---|---|---|---|
| $x^\alpha$ | $C^\alpha t + K^\alpha$ | $C^\alpha t^{2/3} + K^\alpha t$ | $C^\alpha\,\mathrm{e}^{Ht} + K^\alpha\,\mathrm{e}^{-Ht}$ |

The implication of all of this, as shown above, is that a freely falling cloud of objects in an Einstein-de Sitter universe ($R(t) = t^{2/3}$) will feel the expansion of the universe, no matter how small the scale of the cloud. Consequently one might expect the expansion of the universe to imply a spiralling of planetary orbits when a mass such as the Sun is imbedded in an Einstein-de Sitter universe (cf Gautreau 1984). The predictions for the Milne universe (flat spacetime) and steady state universe are also given for comparison, in the first and last columns respectively.

This is a question that has been debated from time to time with differing results (see, e.g., Dicke and Peebles 1964, Noerdlinger and Petrosian 1971, Gautreau 1984). The issue is the relation between a 'lumpy' universe model and an idealised FLRW model. If the FLRW description were indeed applicable to the smallest scales, the above analysis would apply and even the moon's orbit would be affected by the Hubble expansion (note that we are not arguing here whether or not such an effect would be measurable in relation to the effects of all the other forces acting; it is whether there is any Hubble effect at all). However in a lumpy model universe the above description will be modified near a local mass. Instead of the pure Ricci curvature tensor characterised by equation (28), the curvature tensor of the lumpy universe U (which will give the effect of the universe on local physics via equation (27)) will have a Weyl component as well as a perturbed Ricci component, the latter depending on the fitting (or 'gauge') chosen.

In extreme cases the curvature tensor near a massive spherically symmetric body will have only a Weyl component, the Ricci tensor vanishing in some region around the object. This is the approximation introduced by Einstein and Strauss (1945, 1946); a Newtonian version is presented by Callan *et al* (1964). In that case, there is no effect. Inside this region the spacetime takes the form of Schwarzschild solution which is unaffected by any Hubble expansion. The problem with this analysis is that it does not give any guidance as to how to choose the boundaries of this static region (whether to enclose our galaxy, the local group, the local supercluster, etc). The predictions are unclear without a guide as to this choice; which is the issue of on what scale the FLRW metric applies, for on that scale the Hubble effect *must* be felt. Furthermore this analysis will be invalid if there is a uniformly distributed component of matter (e.g., neutrinos) that enters such spherical regions and gives a Ricci contribution to the curvature tensor there, thereby connecting the interior of the region to the expansion of the universe (Noerdlinger and Petrosian 1971). Depending on the nature of clustering, one can obtain any effect between the extreme Ricci-only (full Hubble effect) and Weyl-only (zero Hubble effect) cases.

Regarded from the averaging viewpoint, the issue is one familiar from investigations of how light propagates in lumpy universes (see, e.g., Bertotti 1966, Dyer and Roeder 1981); it is how our fitting procedure leads to an averaging that smooths out the local Weyl effects (in a spacetime where locally the Ricci tensor is zero but the Weyl tensor non-zero) to produce an effective large-scale situation where the converse is true (the Weyl tensor vanishes and the Ricci tensor non-zero). It particularly raises the issue of what is the length scale on which we can expect this averaging to take place. The issue may, for example, be material to the question of galaxy formation (Noerdlinger and Petrosian 1971).

## 6. Conclusion

The lumpy nature of the real universe implies the need for a proper examination of the relation between a 'realistic' universe model U and an idealised FLRW model U′. This paper looks at that issue by considering how, given a lumpy universe model U, one can find a 'best-fit' of a FLRW universe U′ to U.

An adequate fitting procedure will be based on observational criteria, attempting to use all the relevant data available in a coherent way. It will not only determine best-fit parameters for the Robertson–Walker universe U′ but also

(*a*) give a spacetime fit between U′ and the 'lumpy' universe U;

(*b*) characterise the goodness-of-fit achieved, and specify criteria of acceptable fit (under what circumstances would we reject all FLRW models for U);

(*c*) in a well defined way represent how the FLRW model U′ is an average of the lumpy universe U;

(*d*) determine length scales on which the FLRW model describes the universe U well, particularly stating (i) a scale $L_1$ of uniformity (if we average U′ over that length scale, we obtain a FLRW model U to a good approximation) and (ii) a criterion to determine the length scale $L_2$ above which the Hubble law will apply to local physical systems.

It turns out that in practice the fitting problem really has two parts: finding an optimal correspondence between *any* FLRW model and the lumpy universe (deciding which points should be taken to correspond with which points), and then finding the best-fit parameters to describe the FLRW universe given this correspondence, i.e. finding which *particular* FLRW model is optimal. Thus we can regard the fitting problem as consisting of a correspondence problem and an optimal parameter problem.

We do not claim that the null data fitting approach explored in this paper is necessarily the best one. It goes quite a way towards what is needed, but needs development; in particular its implications need exploration, on the one hand tying in this approach to the use of data (element abundances, ages, correlation functions, and above all the microwave background radiation) not directly tied to the spacetime geometry on the past null cone, and on the other examining its relation to perturbation studies. Such further investigation may confirm the usefulness of the approach or suggest alternative ways of developing the theme.

Some alternative possible approaches are suggested in the appendix, each throwing up interesting questions, but all less well related to null cone observations than the null data approach. However it may be that one of them will be better able to use the data that is not direct null cone data. Further work is needed to resolve this question. A particular issue is the relation of this work to the standard perturbation approach in which a perturbation $\delta U$ is added to a FLRW model $U^*$ to give a 'lumpy' universe $U^\dagger$; and then observations are analysed by examining $U^\dagger$. That is, one starts with a FLRW model and ends up with a more realistic model. Our standpoint is the converse. We consider starting with the lumpy universe U, and finding a FLRW model $U'$ and how it best fits U. It is far from clear that these processes commute, i.e. if our best-fit procedure is applied to $U^\dagger$, we will probably not arrive back at $U^*$. This is the 'gauge' problem in perturbation theory: the FLRW model can be fitted in various ways, and the resultant predictions depend on this fit. To first order one can work with gauge invariant quantities, but this is not possible to higher orders. Thus one then has to choose a particular gauge to determine physical results; and the best way to do this is a major aspect of the fitting problem.

Our final comment is that even if the present approach does not succeed as we hope it will, the problem posed is a significant one and we hope this paper will stimulate further study of the questions raised.

## Acknowledgments

We should like to acknowledge helpful discussions with many people, particularly B G Schmidt, J Ehlers, J Winicour, B Bertotti and M J Rees made comments that assisted in this development. This work was initiated at the Max Planck Institute of Astrophysics, Munich; GFRE would like to thank them for their hospitality. We thank the SERC for financial support and Queen Mary College for hospitality.

## Appendix. Alternative approaches

In this appendix, we briefly consider the alternative ways of implementing the 'best-fit' idea mentioned in the introduction.

*A1. The space of spacetimes*

In this approach, in essence U and U′ are represented by points P, P′ respectively in the space of spacetimes S, the point P being constrained to lie in the surface F in S of FLRW spacetimes. Our problem is simply to define a suitable distance function on S; then given U, we choose U′ so as to minimise the distance in S between the given point P and the point P′ on F.

Unfortunately many problems intrude. Firstly, there is no natural positive definite metric defining a distance function on S. Secondly, many different representations $g_{ab}(x^j)$ of the same spacetime can be given because of the arbitrariness of coordinates allowed; indeed to fully describe the manifold M we need in principle a complete atlas of charts $\{x^j\}$, but in practice we will have available a representation in terms of a single coordinate system $\{x^j\}$. Therefore in order to tell if two different points in S represent the same spacetime or not we need to factor out this coordinate freedom (which is the group of diffeomorphisms of M). Quite apart from the problems of doing this in practice, problems arise even in principle because conical singularities occur in the factor space S′. Thirdly, each point in S or S′ represents an entire spacetime (with geometry determined by the metric components $g_{ab}(x^j)$ as $x^j$ varies over the four-dimensional manifold M); this means that the fitting envisaged would involve some kind of comparison of the metric tensors at all pairs of points in the two spacetimes, by means of some integration over all spacetime points. This cannot easily be related to a possible observational procedure on the past null cone. Finally we actually need to do more than fit the metrics: we need to fit the matter distributions in the two spaces by fitting at least the 4-velocities $u^a$, $u'^a$ and energy densities $\mu$, $\mu'$ as well, so we need an extended version of the space of spacetimes that also includes this information.

The second problem can be resolved by stating our aim as being to find the best fit in a specific (operationally defined) coordinate system rather than general coordinates; then we do not need to factor out the coordinate freedom. However the first and the last two problems remain; nevertheless this approach gives a useful concept of what we are trying to do.

*A2. The initial data approach*

The first proposal does not take into account the dynamics of general relativity. One can do so by going to the phase space of a cosmological model in general relativity. The idea is similar to the previous one, but instead of considering the space of spacetimes we consider the space S* of initial data for a spacetime, where the initial data are given on a spacelike surface Σ. These initial data consist of the metric ${}^3g_{ab}$ of Σ, its second fundamental form $\pi^{ab}$, and the matter energy density $\mu$ and 3-momentum $q^a$ relative to Σ; thus each point Q in S* will correspond to a specification of all of these quantities at each point on a spacelike 3-surface Σ, chosen so as to satisfy the Einstein constraints on Σ.

The problems of the previous approach remain. The comparison problem is marginally ameliorated, for one only has to compare quantities at all points on a three-dimensional manifold Σ rather than a four-dimensional spacetime M. However, the second problem is exacerbated, for again to distinguish distinct spacetimes one has to factor out the diffeomorphisms of Σ and the result is not a manifold, but additionally quite different data can represent the same cosmology U (simply choose two different spacelike slicings Σ of U and for example the 3-space metrics ${}^3g_{ab}$ can be quite different).

One may query if this indeterminacy is significant: if we could determine the required fit of 3-metrics, the four-dimensional fit would be determined, even if not explicitly known. In any case we are really involved here with the Hamiltonian structure of general relativity, and the symplectic form of that structure gives a way of comparing metrics on two different spacelike surfaces to see if they represent slicings of the same spacetime. Thus this problem may not be insuperable; indeed this symplectic form can be used to give a promising measure on the set of cosmological models (Stewart *et al* 1986). Nevertheless it seems difficult to relate this approach directly to a suitable observationally-based fitting procedure of the type we desire. However the adaptation of such an 'initial data' idea to null surfaces is possible (see § A7).

### *A3. Near-equivalence*

Clearly we are asking a question that is close to the classic equivalence problem of general relativity, reviewed by Ehlers (1981) and Aman *et al* (1985). A direct approach by comparison of curvature invariants of two spacetimes is problematic because the metric tensors are indefinite; however equivalence of the curvature tensors and their covariant derivatives, evaluated in orthonormal frames, is sufficient to prove the equivalence of two spacetimes.

One should carefully distinguish here the analytic case from the $C^\infty$ case. If the spacetime is analytic, we can consider quantities at a single point p in M and p′ in M′. If the curvature tensor and its $s$th covariant derivatives for all $s = 1, 2, 3, \ldots$ are equivalent at these two points (i.e. orthonormal frames $e$, $e'$ can be found at p, p′ so that all their components are identical) then the spacetimes are locally equivalent. To determine if this is so or not is a purely algebraic problem.

If the spacetime is $C^\infty$, this condition (at a single pair of points) is not sufficient to show equivalence. However in this case if the curvature tensor and its $s$th covariant derivatives for all $s = 1, 2, 3, \ldots, k$ (evaluated in suitable orthonormal frames in each spacetime) are equivalent in open neighbourhoods of p, p′, then this shows equivalence of the spacetimes in those open neighbourhoods, where $k$ is a number depending on the algebraic structure of the curvature tensor, which is not larger than $n(n+1)/2$ if $n$ is the dimension of the spacetimes. Thus one has only to consider a finite number of derivatives, but their functional dependence has to be compared on two open sets—no longer an algebraic problem at a point. However before the functional issue is tackled (which in a sense depends on searching all possible choices of identification of points in U, U′ as well as all orthonormal frames at each pair of points p, p′) one can check necessary conditions for identification: namely that both spacetimes have the same 'generalised Petrov type' (Karlhede 1980), i.e. the same algebraic structure for the curvature tensor and its covariant derivatives. The practical issue then becomes primarily one of determining how many of the components of the curvature tensor and its first $s$ covariant derivatives are functionally independent in each of the spaces to be compared, for each value of $s$; and this can be resolved by algebraic means at each point (by examining the rank of matrices of partial derivatives of tensor components in each of the spacetimes). A formal statement of the procedures is most conveniently made in terms of the bundle of orthonormal frames over each spacetime, but practical calculations are most conveniently carried out in terms of specific orthonormal tetrads in each spacetime, which effectively handles the problem of choosing appropriately related orthonormal tetrads in the two spaces. The relevant calculations can largely be carried out by computer algebraic systems, and this makes the approach a practical way of tackling the equivalence problem (Karlhede 1980, Aman *et al* 1985).

Given the success of this approach to the equivalence problem, one can ask if an extension of these methods might not be used to determine if two cosmologies are almost equivalent. Various problems arise. Firstly, there is the question already mentioned of how to choose which pairs of points p, p′ to try to identify in the two cosmologies p, p′. Secondly, the idea of 'almost equivalence' is easiest to express in the context of the frame bundle which is the theoretical base for this work (roughly 'there exist tetrads in which the components of the curvature tensor and its covariant derivatives up to the $k$th are nearly the same in both open sets') but is more complex in the context of specific choices of tetrad canonically related to particular Petrov types (e.g. given a tetrad adapted to a type-N Weyl tensor, which Weyl tensors of type III are 'nearby'?). This kind of question can be answered by applying methods introduced by Arnold to determine stable matrix forms, for example to derive an extended Petrov classification of the Weyl tensor (see McCarthy and Ellis 1987). Such an extended algebraic classification of the curvature tensor and its derivatives should make possible determination of 'almost equivalence' of spacetimes by extending the methods used to examine exact equivalence.

The problem is simplified in the cosmological context by the existence of a preferred 4-velocity $u^a$, for this reduces the set of tetrads we have to examine (choosing this vector as the timelike tetrad vector in each spacetime, we only have to allow small boosts plus arbitrary rotations in the tetrad freedom in each spacetime). If we adopt the usual perfect fluid description for matter, then the preferred vector field $u^a$ is a timelike eigenvector of the Ricci tensor which is isotropic in the 3-space orthogonal to $u^a$. Thus when $u^a$ and the metric tensor $g_{ab}$ are known the Ricci tensor is determined by two scalar invariants, namely $\mu$ (the energy density) and $p$ (the pressure). Furthermore the Weyl tensor is then determined by two trace-free symmetric tensors $E_{ab}$ and $H_{ab}$ (see, e.g., Ellis 1971). Thus the classification problem is easier in the cosmological context than in general, and the question of 'almost equivalence' might be quite tractable. This is a promising approach which could be of interest, but its relation to possible observational procedures is obscure.

### *A4. Normal coordinates*

A 'rough and ready' approach to equivalence is based on the fact that in the analytic case we can examine the curvature near any spacetime point by using normal coordinates about that point. The deviation of the metric tensor components in these coordinates from their Minkowski values directly reflects the spacetime curvature (see, e.g., the discussion of 'locally geodesic coordinates' by Ehlers (1973)). Thus we can compare different spacetimes locally by writing each of them in such coordinate systems centred on points p, p′ and then making a direct comparison of the metric components up to some required order of accuracy (this is really just the power series approach to the analytic case, mentioned in the previous section); in this way we can characterise the idea of an almost equivalence of the spacetimes locally. As before, a significant problem is how we choose which correspondence of points p, p′ to make in the two spacetimes.

If we relate this local approach to the question of a 'best fit' to astronomical observations, we will essentially end up with an adaptation of the Kristian and Sachs (1966) observational cosmology analysis on the past null cone. However if we are going to go this route, it would presumably be better to carry out the analysis in a non-local way, thus ending up with the 'null-data' approach to fitting discussed in § 4 of this paper. Nevertheless other useful adaptations of the 'local' approach on the past light cone might be possible; see, e.g., Penrose (1985).

A normal-coordinate approach to fitting is closely related to the nature of local physics in a lumpy universe, as discussed in § 5.

### *A5. Gauge choice*

As mentioned in § 1.5 the choice of gauge in describing perturbations of a FLRW universe is essentially a question of fitting a FLRW universe U′ to a lumpy universe U (Sachs and Wolfe 1967, Stewart and Walker 1974, Bardeen 1980). This is illustrated in figure 6: let $\{x^i\}$ be local coordinates in the FLRW universe U, and $\{y^j\}$ local coordinates in the lumpy universe U′. Each of these coordinates are maps from the spacetime into $\boldsymbol{R}^4$ that can be chosen arbitrarily (because we can choose coordinates freely in each spacetime). A choice of gauge is a 1-1 correspondence $\psi$ between U and U′: $p' \in U' \to p = \psi(p') \in U$. This is conveniently expressed in terms of functions $f^j$ stating the relation between the coordinates of p and p′: $y^j(p) = f^j(x^i(p'))$. A gauge transformation is a change in the correspondence $\psi$. This could be expressed in terms of change in the functions $f^j$ (keeping the coordinates $x^i$, $y^j$ fixed in each spacetime), but that is not the usual procedure. Rather it is usually assumed that the correspondence is defined by assigning the same coordinates to p and p′, i.e.

$$y^j(p) = f^j(x^i(p')) = x^i(p'). \tag{A1}$$

In this case, a gauge transformation is represented (Bardeen 1980, p 1882) by specifying a change in the coordinates $y^j$ in U, keeping fixed the coordinates $x^i$ in U′ and the coordinate relation (A1). Most presentations thus talk about gauge transformations as coordinate transformations in U′, without emphasising that the essence of a gauge choice is a choice of point correspondence between the two spacetimes.

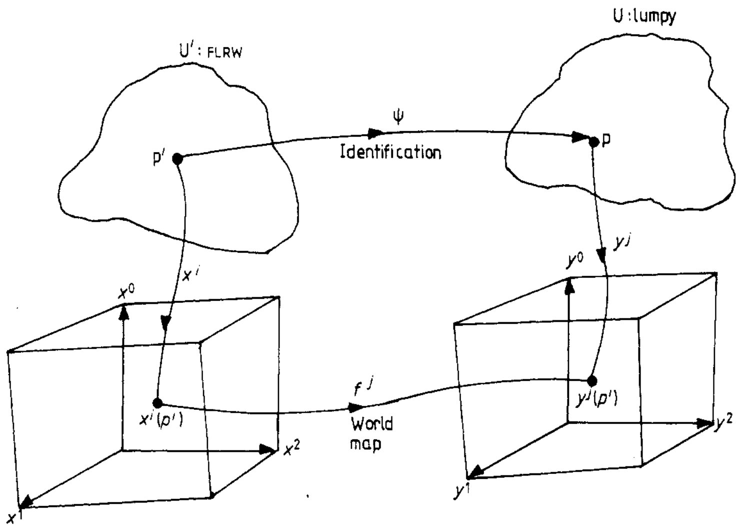


**Figure 6.** Choice of a gauge in a perturbed universe model U is equivalent to choosing a point identification between the FLRW model U′ and the lumpy universe U. Given a choice of local coordinates in U′ and in U, the identification can be expressed in terms of a relation between these coordinates. This is usually taken to be the identity, i.e. the coordinates of the corresponding points are taken to be the same.

A specific gauge can be characterised in terms of a choice of 'hypersurfaces of simultaneity in the physical spacetime' (Bardeen 1980), which is actually the choice of a mapping of the FLRW surfaces ($t$ = constant) in U′ into the lumpy universe model U. The choices for these surfaces investigated by Bardeen are comoving hypersurfaces, zero-shear hypersurfaces, and uniform-Hubble-constant hypersurfaces. Alternatively in a 'synchronous gauge' a choice is made for one of these surfaces at an initial time, and the later surfaces are then determined non-locally by solution of differential equations. The corresponding 'best-fit' scheme would proceed by using one of these prescriptions to set up a point correspondence between spacetimes, and then look for the parameters $H_0$, $q_0$ that gave a best fit between U and U′.

### *A6. Averaging*

Best-fitting by averaging over a suitable length scale is outlined in § 1.3 and discussed in § 2. The major problem of arbitrariness of the spatial surfaces on which averaging is carried out is resolved if the approach is adapted to averaging on a past null cone, cf § 4.6.

### *A7. Null data*

Best-fitting on the basis of data specified on a null cone is the main theme of this paper, see § 4 above. It can be regarded as a way of realising the space of spacetimes approach (§ A1) by adapting the initial value approach (§ A2) to the past null cone. In effect this leads to a specific choice of gauge in perturbation theory (§ A5) based on the past null cone, but the implications of this choice have yet to be worked out. The null data approach does not directly relate to the near-equivalence approach (§ A3), but is related to the normal coordinate approach (§ A4) through the Kristian and Sachs (1967) paper, although again this has not been worked out in detail. The averaging approach (§ A6) can be adapted to the past null cone, and indeed observationally viewed this is the proper setting for that approach.

## References


Aman J E, d'Inverno R A, Joly G C and MacCallum M A H 1985 *Eurocal '85* vol 2 (*Lecture Notes in Computer Science* **204**) ed B F Caviness (Berlin: Springer) p 89

Bardeen J M 1980 *Phys. Rev.* D **22** 1882

Bertotti B 1966 *Proc. R. Soc.* A **294** 195

Bonnor W B and Ellis G F R 1986 *Mon. Not. R. Astron. Soc.* **218** 605

Bondi H 1960 *Cosmology* (Cambridge: Cambridge University Press)

Callan C, Dicke R H and Peebles P J E 1964 *Am. J. Phys.* **33** 105

Carfora M and Marzuoli A 1984 *Phys. Rev. Lett.* **53** 2445

D'Eath P 1976 *Ann. Phys., NY* **98** 237

Dicke R H and Peebles P J E 1964 *Phys. Rev. Lett.* **12** 435

Dyer C C and Roeder R C 1981 *Gen. Rel. Grav.* **13** 1157

Ehlers J 1973 *Relativity, Astrophysics and Cosmology* ed W Israel (Dordrecht: Reidel)

—— 1981 *E B Christoffel* ed P L Butzer and F Feher (Basel: Birkhauser)

Einstein A and Straus E G 1945 *Rev. Mod. Phys.* **17** 120

—— 1946 *Rev. Mod. Phys.* **18** 148

Ellis G F R 1971 *General Relativity and Cosmology, Enrico Fermi Summer School Course XLVII* ed R K Sachs (New York: Academic)

—— 1975 *Q. J. R. Astron. Soc.* **16** 245

—— 1980 *Ann. NY Acad. Sci.* **336** 130
—— 1984 *General Relativity and Gravitation, GR10 Conf. Rep.* ed B Bertotti (Dordrecht: Reidel) p 215
Ellis G F R and Baldwin J 1984 *Mon. Not. R. Astron. Soc.* **206** 377
Ellis G F R, Nel S D, Maartens R, Stoeger W R and Whitman A P 1985 *Phys. Rep.* **124** 316
Ellis G F R and Schreiber G 1986 *Phys. Lett.* **115A** 97
Gautreau R 1984 *Phys. Rev.* D **29** 198
Gott J R, Gunn J E, Schramm D M and Tinsley B M 1974 *Astrophys. J.* **194** 543
Hoyle F 1960 *Rend. Scuola Int. di Fis. E Fermi XX Course* (New York: Academic)
Karlhede A 1980 *Gen. Rel. Grav.* **12** 693
King-Hele D G and Walker D M C 1986 *Planet. Space Sci.* **34** 183
Kristian J and Sachs R K 1966 *Astrophys. J.* **143** 379
Matzner R A 1968 *J. Math. Phys.* **9** 1957
McCarthy P and Ellis G F R 1987 *Ann. Phys., NY* to be published
Misner C W, Thorne K S and Wheeler J A 1972 *Gravitation* (New York: Freeman)
Noerdlinger P D and Petrosian V 1971 *Astrophys. J.* **168** 1
Penrose R 1985 *Galaxies, Axisymmetric Systems and Relativity* ed M A H MacCallum (Cambridge: Cambridge University Press)
Rowan-Robinson M 1985 *The Cosmological Distance Scale* (New York: Freeman)
Sachs R K and Wolfe A 1967 *Astrophys. J.* **147** 73
Sandage A 1968 *Observatory* **88** 91
Sandage A and Tammann G A 1986 *Inner Space/Outer Space* ed E W Kolb, M S Turner, D Lindley, K Olive and D Seckel (Chicago: University of Chicago Press)
Stebbins A and Silk J 1986 *Astrophys. J.* **300** 1
Stewart J M, Gibbons G W and Hawking S W 1986 *Preprint, Cambridge University*
Stewart J M and Walker M 1974 *Mon. Not. R. Astron. Soc.* A **341** 49
Stoeger W, Ellis G F R and Hellaby C 1987 *Mon. Not. R. Astron. Soc.* **226** 373
van den Bergh S and Pritchett R 1986 *Publ. Astron. Soc. Pac.* **98** 110
Wagoner R V 1977 *Astrophys. J.* **214** L5
Weinberg S 1972 *Gravitation and Cosmology* (New York: Wiley)
Yahil A, Walker D and Rowan-Robinson M 1986 *Astrophys. J.* **301** L1