Defining the frame of minimum nonlinear Hubble expansion variation

James H. McKay⋆ & David L. Wiltshire†
Department of Physics & Astronomy, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand

ABSTRACT

We characterize a cosmic rest frame in which the monopole variation of the spherically averaged nonlinear Hubble expansion is most uniform, under arbitrary local Lorentz boosts of the central observer. Using the COMPOSITE sample of 4534 galaxies, we identify a degenerate set of candidate minimum nonlinear variation frames, which includes the rest frame of the Local Group (LG) of galaxies, but excludes the standard Cosmic Microwave Background (CMB) frame. Candidate rest frames defined by a boost from the LG frame close to the plane of the galaxy have a statistical likelihood similar to the LG frame. This may result from a lack of constraining data in the Zone of Avoidance. We extend our analysis to the *Cosmicflows*-2 (CF2) sample of 8162 galaxies. While the signature of a systematic boost offset between the CMB and LG frame averages is still detected, the spherically averaged nonlinear expansion variation in all rest frames is significantly larger in the CF2 sample than would be reasonably expected. We trace this to the CF2 distances being reported without a correction for inhomogeneous distribution Malmquist bias. Systematic differences in the inclusion of the large SFI++ subsample into the COMPOSITE and CF2 catalogues are analysed. Our results highlight the importance of a careful treatment of Malmquist biases for future peculiar velocities studies, including tests of the hypothesis of Wiltshire et al. (2013) [Phys. Rev. D, 88, 083529] that a significant fraction of the CMB temperature dipole may be nonkinematic in origin.

Key words: cosmology: observations — cosmology: theory — distance scale

1 INTRODUCTION

Although the Universe is spatially homogeneous in some statistical sense, at the present epoch it exhibits a complex hierarchical structure, with galaxy clusters forming knots, filaments and sheets that thread and surround voids, in a complex cosmic web (Forero–Romero et al. 2009; Bilicki et al. 2014; Einasto 2014). Deviations from homogeneity are conventionally treated in the framework of peculiar velocities, by which the mean redshift, z, and luminosity distance, r, of a galaxy cluster are converted to a peculiar velocity at small redshifts according to

$$v_{\rm pec} = cz - H_0 r \qquad (1)$$

where c is the speed of light and $H_0 = 100\,h\,{\rm km\ sec^{-1}\ Mpc^{-1}}$ is the Hubble constant.

While large galaxy surveys reveal that statistical homogeneity emerges on scales $70 - 100\,h^{-1}{\rm Mpc}$

(Hogg et al. 2005; Scrimgeour et al. 2012) as determined by the two–point galaxy correlation function, in the standard peculiar velocities framework relatively large bulk flows are seen on scales larger than this relative to our own location. The amplitude of such flows, and their consistency with the standard Lambda Cold Dark Matter (ΛCDM) cosmology, are a matter of ongoing debate (Watkins, Feldman & Hudson 2009; Kashlinsky et al. 2010; Nusser & Davis 2011; Turnbull et al. 2012; Lavaux et al. 2012; Ma & Scott 2013; Atrio-Barandela 2013; Ade et al. 2014a; Hoffman, Courtois & Tully 2015; Carrick et al. 2015). While it may be possible to achieve convergence of such bulk flows to the CMB frame within current uncertainties based on cosmic variance and N-body simulations in the ΛCDM model, given the deep mystery associated with the nature of dark energy, it is important that alternative cosmological models are investigated. Such alternatives (Buchert 2000, 2008; Wiltshire 2007a,b, 2009) may also demand a treatment of cosmic expansion which includes effects of inhomogeneities that fall outside the peculiar ve-

---

⋆ E-mail: j.mckay14@imperial.ac.uk
† E-mail: david.wiltshire@canterbury.ac.nz

locity framework, with testable observational consequences (Li & Schwarz 2008; Wiltshire et al. 2013).

The peculiar velocity framework makes a strong geometrical assumption over and above what is demanded by general relativity. In particular, the quantity $v_{\rm pec}$ defined by (1) only has the physical characteristics of a velocity if one implicitly assumes the spatial geometry *on all scales* larger than those of bound systems is exactly described by a homogeneous isotropic Friedmann-Lemaître-Robertson-Walker (FLRW) model with a single cosmic scale factor, $a(t)$, whose derivative defines a single global Hubble constant, $H_0 = \dot{a}/a|_{t_0}$. Deviations from the uniform expansion are then ascribed to Lorentz boosts of each galaxy cluster with respect to the spatial hypersurfaces of average homogeneity, at each cluster location.

It is a consequence of general relativity, however, that inhomogeneous matter distributions generally give rise to a differential expansion of space that cannot be reduced to a single uniform expansion plus local boosts. This is a feature of general exact solutions to the cosmological Einstein equations, such as the Lemaître–Tolman–Bondi (LTB) (Lemaître 1933; Tolman 1934; Bondi 1947) and Szekeres (1975) models. Any definition of the expansion rate in such models depends on the spatial scale relative to that of the inhomogeneities. Although one can define scale dependent Hubble parameters for specific exact solutions – for example, given the spherical symmetry of the LTB model – the actual cosmic web is sufficiently complex that in reality one must deal with spatial or null cone averages in general relativity.

In recent work Wiltshire et al. (2013) examined the variation[1] of the Hubble expansion from a fresh perspective, by generalizing the earlier approaches of Li & Schwarz (2008) and McClure & Dyer (2007). In particular, given that there is a notion of statistical homogeneity on large ($\gg 100\,h^{-1}$Mpc) scales, then an average expansion law characterized by a single asymptotic Hubble constant, $\bar{H}_0$, is applicable on such scales. However, from the first principles of general relativity one should make no geometrical assumptions about cosmic expansion below the statistical homogeneity scale, where cosmic expansion is nonlinear. One can nonetheless perform radial and angular averages of the distance versus redshift of a large sample of galaxies in spherical shells, and compare the results with the asymptotic Hubble constant in order to quantify the nonlinear variation of the Hubble expansion.

Wiltshire et al. (2013) conducted such an analysis on the COMPOSITE sample of 4534 cluster, group and galaxy distances (Watkins, Feldman & Hudson 2009; Feldman, Watkins & Hudson 2010), with the following results:

- A linear Hubble law with a spherically averaged Hubble constant, H_s, which is statistically indistinguishable from the asymptotic Hubble constant, $\bar{H}_0$, is found to emerge in independent radial shells with mean distances in the range $\bar{r}_s > 70\,h^{-1}$Mpc.

- On scales $r \lesssim 65\,h^{-1}$Mpc the spherically averaged value, H_s, in independent shells is greater than the asymptotic value, $\bar{H}_0$. However, the difference is significantly larger in the standard rest frame of the CMB radiation when compared to either the rest frame of the Local Group (LG) of galaxies or the Local Sheet (LS). In other words, the spherically averaged Hubble expansion is more uniform in the LG rest frame than in the CMB rest frame, with very strong Bayesian evidence, $\ln B \gg 5$. The uniformity of expansion in the LG and LS frames is statistically indistinguishable.

- By a variety of angular tests, the residual variation of the spherical (monopole) Hubble expansion is found to be correlated with structures in the range $32 - 62\,h^{-1}$Mpc, which give a Hubble expansion dipole with a markedly different character in the CMB and LG frames.

- A skymap of angular variation of the Hubble expansion in the LG frame, constructed by Gaussian window averaging (McClure & Dyer 2007), has a very strong dipole. The angular expansion skymap has a correlation coefficient of -0.92 with the residual CMB temperature dipole in the LG frame.

- On scales $\gtrsim 80\,h^{-1}$Mpc the magnitude of the dipole variation in the Hubble expansion is less in the LG frame than in the CMB frame. While the LG frame dipole is statistically consistent with zero in most individual outer shells, by contrast the CMB frame dipole magnitude rises to a residual level. Consequently the large bulk flow reported by Watkins et al. (2009) may be partly due to a choice of rest frame.

The first of the results above is consistent with other observations which find that a notion of statistical homogeneity emerges at scales of order $70 - 100\,h^{-1}$Mpc (Hogg et al. 2005; Scrimgeour et al. 2012). Furthermore, the fact that $H_s > \bar{H}_0$ on the $\lesssim 65\,h^{-1}$Mpc scales on which the Hubble expansion is nonlinear agrees well with the observation that the largest typical[2] structures in the late epoch Universe are voids of diameter $30\,h^{-1}$Mpc (Hoyle & Vogeley 2002, 2004). Pan et al. (2011) found that voids occupy 62% of the volume studied in the 7th release of the Sloan Digital Sky Survey (Abazajian et al. 2009). If one constructs averages in spherical shells, as in Fig. 1, then once the shells are a few times larger than the diameter of the largest typical nonlinear structures, a well-defined asymptotic average, $\bar{H}_0$, is obtained which does not change when shells are further enlarged. When shells are $1 - 2$ times the diameter of the typical nonlinear voids, however, a variation in expansion rate is seen and since the faster expanding voids dominate by volume then the average, H_s, is increased relative to $\bar{H}_0$.

While the results just discussed may not be surprising to those familiar with the statistics of the cosmic web, the remaining results of Wiltshire et al. (2013) listed above are not at all expected in the conventional approach to peculiar velocities in observational cosmology. In particular, the cosmic rest frame should be the one in which the variation of cosmic expansion is a minimum in some sense. The results of Wiltshire et al. (2013) show that such a frame of

---

[1] Wiltshire et al. (2013) used the terminology "Hubble flow variance", where variance was understood in the same loose sense as in the terminology "cosmic variance". Here we use the more generic word "variation", to avoid any confusion with the square of a standard deviation.

[2] Larger structures such as the $320\,h^{-1}$Mpc long Sloan Great Wall (Gott et al. 2005) and the $350\,h^{-1}$Mpc long Large Quasar Group (Clowes et al. 2013) are known, but it is arguable that these are not typical.
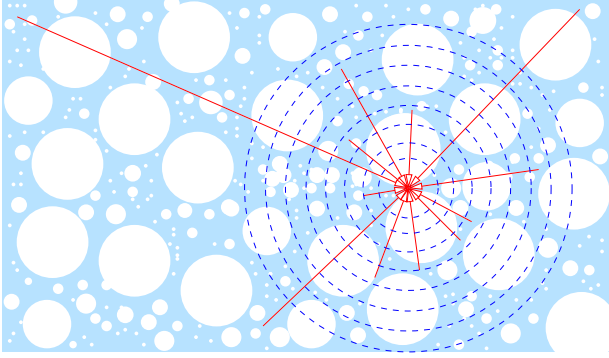
Figure 1. Schematic diagram of spherical averaging (from Wiltshire et al. (2013)). The universe is an ensemble of filaments, walls and voids, the largest *typical* nonlinear structures being voids of diameter $\sim 30\,h^{-1}$Mpc (Hoyle & Vogeley 2002, 2004; Pan et al. 2011). If one averages cz/r in spherical shells (dotted lines) about any point then once the shells are a few times larger than the typical nonlinear structures, an average Hubble law with small statistical scatter is obtained (e.g., for the longest null geodesics depicted by lines converging to the central observer). However, there are considerable deviations for shells on scales comparable to the typical nonlinear structures. This approach makes no assumptions about spatial geometry of the universe on the scales in question, and thus is model independent.

minimum nonlinear Hubble expansion variation is not the conventional CMB frame.

One important consequence is that a substantial fraction of the CMB dipole usually attributed to a local boost of the Local Group at 635 ± 38 km sec^{-1} (Tully et al. 2008) in the direction $(\ell, b) = (276.4°, 29.3°)$ would be due to a nonkinematic differential expansion of space. It has been independently shown that the radio galaxy number count dipole appears to be nonkinematic, at the 99.5% confidence level, using the NRAO VLA Sky Survey (Rubart & Schwarz 2013).

Wiltshire et al. (2013) estimate that a 0.5% anisotropy in the distance–redshift relation on $\lesssim 65\,h^{-1}$Mpc scales would be required to achieve the observed properties of the CMB dipole. Furthermore, using ray–tracing simulations on an exact Szekeres solution of Einstein's equations which exhibits differential expansion on such scales, it is found that it is possible to produce a Hubble expansion dipole which is very close to that of the COMPOSITE sample, while simultaneously generating CMB anisotropies that both account for the residual temperature dipole in the LG frame and are consistent with the amplitude of higher order multipoles (Bolejko, Nazer & Wiltshire 2015).

This may potentially provide an explanation for the angular scale dependence observed in attempts to measure the boost to the CMB frame using the effects of special relativistic aberration and modulation on the CMB anisotropy spectrum (Aghanim et al. 2014). Since subtraction of a kinematic dipole is a key step in the map making process, this may also have a direct impact on various large angle anomalies in the CMB anisotropy spectrum (Ade et al. 2014b). Furthermore, given that a transformation to the CMB frame is a step that is taken in many observational procedures without question, the possibility that a substantial fraction

of the CMB dipole may be nonkinematic could have a large impact in other areas of observational cosmology.

Given the potential importance of such a result, it is important to try to characterize the frame of minimum nonlinear Hubble expansion variation in purely observational terms. Wiltshire et al. (2013) compared the LG and LS frames with that of the CMB, motivated by the starting point that a frame close to the LG frame would be the natural standard of rest according to the "Cosmological Equivalence Principle" (Wiltshire 2008) which underlies the approach of Wiltshire (2007a,b, 2009) to the averaging problem in inhomogeneous cosmology (Ellis 1984; Buchert 2000, 2008; Wiltshire 2011; Nazer & Wiltshire 2015).

This paper asks the question – if we make no assumptions about our own standard of rest, but make arbitrary boosts at our own location – can we determine in a model–independent fashion whether there is a rest frame in which the nonlinear variation of the Hubble expansion is minimized? We will consider all scales over which cosmic expansion is nonlinear and transitions to linearity, and the question of how such a transition impacts the characterization of a minimum variation frame.

While our principal results are determined from an analysis of the COMPOSITE sample (Watkins et al. 2009; Feldman et al. 2010), we have also considered the recently published *Cosmicflows*-2 (CF2) sample of Tully et al. (2013). We find that the issue of the different treatments of Malmquist bias in the two datasets at present prevents as detailed analysis as we perform for the COMPOSITE sample. However, our investigation highlights how the implicit assumption of a FLRW expansion law below the scale of statistical homogeneity via (1) subtly influences the manner in which such biases are treated in practice, and raises concerns about remaining distribution biases in the CF2 distances.

2 METHODOLOGY

In this paper we will use the methodology of Wiltshire et al. (2013), who considered both radial and angular averages of the Hubble expansion in the COMPOSITE sample, adapting techniques that Li & Schwarz (2008) and McClure & Dyer (2007) had previously applied to radial and angular averages respectively in the much smaller Hubble Telescope Key Project dataset (Freedman et al. 2001). We are primarily interested in the radial spherical averages, as it was by this method that Wiltshire et al. (2013) found the most decisive evidence that the Hubble expansion is significantly more uniform in the LG or LS frames, as compared to the standard CMB rest frame.

We begin by repeating the analysis of Wiltshire et al. (2013), but without assuming that our own standard of rest is known. In particular, we will make arbitrary local boosts of the central observer, and perform numerical studies on the COMPOSITE sample to determine what cosmic rest frame(s) are singled out by minimizing the variation in the spherically averaged nonlinear Hubble expansion.

An arbitrary boost of the central observer has the effect of transforming the inferred redshift of individual sources we observe from values, z_i, in any one rest frame to new values given by

$$cz_i' = cz_i + v \cos \phi_i \qquad (2)$$

where v is the boost magnitude at our origin, and ϕ_i is the angle between each data point and the boost direction. We stress that in our nonlinear framework the boost velocity v of the central observer is a distinct quantity from the peculiar velocities of individual sources[3], $v_{\rm pec}$, in the linear peculiar velocities framework (1).

Following Wiltshire et al. (2013), we determine the best fit linear Hubble law – even in the regime in which the average expansion is nonlinear – by standard linear regression in independent radial shells, minimizing the quantity

$$\chi_s^2 = \sum_{i=1}^{N_s} \left[\sigma_i^{-1}(r_i - cz_i/H) \right]^2 , \qquad (3)$$

with respect to H, where z_i, r_i and σ_i are respectively the redshift, distance and distance uncertainty of each object, and N_s is the number of data points in a given shell, s. This method, with distance as function of redshift, is chosen because all uncertainties in the COMPOSITE sample have been included as distance uncertainties. Any corrections that would be required due to noise arising from peculiar motion within galaxy clusters have been accounted for by assigning a distance and uncertainty to the cluster itself rather than individual galaxies (Watkins et al. 2009). The values of z_i are taken to be exact, whereas the distance uncertainties are large, approximately 15% for most COMPOSITE sample objects. Fortunately, the large number of data points within each shell nonetheless lead to statistically significant results.

The value of the Hubble constant in the sth shell is then determined as

$$H_s = \left(\sum_{i=1}^{N_s} \frac{(cz_i)^2}{\sigma_i^2}\right) \left(\sum_{i=1}^{N_s} \frac{cz_i r_i}{\sigma_i^2}\right)^{-1} \qquad (4)$$

while its weighted average luminosity distance is

$$\bar{r}_s = \left(\sum_{i=1}^{N_s} \frac{r_i}{\sigma_i^2}\right) \left(\sum_{i=1}^{N_s} \frac{1}{\sigma_i^2}\right)^{-1} . \qquad (5)$$

The radial averages are computed in two different shell configurations of 11 shells, for luminosity distances with $r_s < r \leqslant r_{s+1}$. Both configurations use shells of width $12.5 \, h^{-1}$Mpc in most cases, but start from a different innermost shell cutoff of either $2 \, h^{-1}$Mpc or $6.25 \, h^{-1}$Mpc. The two shell configurations are labelled using unprimed and primed integers respectively, as given in Table 1 of Wiltshire et al. (2013). Given less data at large distances, the shells 10 and $10'$ are taken to be wider than the rest so as to include approximately the same number of data points as the inner shells. They both have an outer cutoff at $156.25 \, h^{-1}$Mpc. The outermost shell 11 has an outer bound given by the largest distance in the sample, and is the same in both configurations, to provide an anchor for the asymptotic value of the Hubble constant, $\bar{H}_0$.

The uncertainty in (4) is taken as

$$\bar{\sigma}_s = \sqrt{\bar{\sigma}_{0s}^2 + \bar{\sigma}_{1s}^2} \qquad (6)$$

---

[3] In particular, our minimum nonlinear Hubble expansion variation frame should not be confused with the terminology of Watkins et al. (2009); Feldman et al. (2010); Agarwal et al. (2012), who use the term "minimum variance" with respect to weights of peculiar velocities in bulk flow studies.

where

$$\bar{\sigma}_{0s} = H_s \frac{\sigma_0}{\bar{r}_s} \qquad (7)$$

and

$$\bar{\sigma}_{1s} = \left(\sum_{i=1}^{N_s} \frac{(cz_i)^2}{\sigma_i^2}\right)^{3/2} \left(\sum_{i=1}^{N_s} \frac{cz_i r_i}{\sigma_i^2}\right)^{-2} , \qquad (8)$$

A zero point uncertainty (7) is added in quadrature to the standard uncertainty (8) determined through error propagation in (4), to yield the total uncertainty (6) for the Hubble constant in each shell. The zero point uncertainty arises from the fact that a linear Hubble law must necessarily pass through the origin. However, there is an uncertainty in determining the origin, due to a $20 \, \text{km} \, \text{sec}^{-1}$ uncertainty in the heliocentric peculiar velocity of the LG (and LS) frames (Tully et al. 2008) and a 0.4% uncertainty in the magnitude of the CMB dipole (Fixsen et al. 1996), which combine to give $\sigma_0 = 0.201 \, h^{-1}$Mpc (Wiltshire et al. 2013). For each shell, the weighted zero point uncertainty (7) is the uncertainty in the linear Hubble law for a shell with mean distance $\bar{r}_s$ due to the uncertainty in the origin alone. This uncertainty is more significant for shells with a small mean distance compared to those at a larger distance, due to the shorter lever arm of the linear Hubble law for shells with smaller mean distances.

As a measure of the difference in the local Hubble expansion from its asymptotic value, we use the quantity (Li & Schwarz 2008; Wiltshire et al. 2013)

$$\delta H_s = \frac{\left(H_s - \bar{H}_0\right)}{\bar{H}_0} \qquad (9)$$

where $\bar{H}_0$ is the mean asymptotic value of the Hubble constant. In our case, $\bar{H}_0$, and its uncertainty are calculated from the data points at distances $r > 156.25 \, h^{-1}$Mpc (shell 11 in both unprimed and primed configurations). Wiltshire et al. (2013) choose the distance scale for the outermost shell as one which is statistically reliable for the COMPOSITE data, while including only data at distances larger than the baryon acoustic oscillation scale, the largest scale on which one might expect to see the effects of inhomogeneity on the local Hubble expansion. They verify that a linear Hubble law with a very high goodness of fit is found for $r > 156.25 \, h^{-1}$Mpc, with a Hubble constant $\bar{H}_0 = (100.1 \pm 1.7)h \, \text{km} \, \text{sec}^{-1} \, \text{Mpc}^{-1}$ in the CMB frame or $\bar{H}_0 = (101.0 \pm 1.7)h \, \text{km} \, \text{sec}^{-1} \, \text{Mpc}^{-1}$ in the LG/LS frames, consistent with the normalization $H_0 = 100 \, \text{km} \, \text{sec}^{-1} \, \text{Mpc}^{-1}$ used in the COMPOSITE sample.

3 MONOPOLE EXPANSION VARIATION DUE TO SYSTEMATIC BOOST OFFSETS

Wiltshire et al. (2013) propose that the much larger monopole variation of the linear Hubble law observed in the CMB frame, as compared to the LG/LS frames, may have a systematic origin. This arises from the nonlinear dependence of H_s in (4) on the individual cz_i values, which change when performing a local Lorentz boost to a different frame. By calculating the result of an arbitrary boost on the individual H_s values Wiltshire et al. (2013) obtain an explicit form for this systematic variation.

Consider redshifts, z_i, observed in a frame of reference in which the variation of the spherically averaged Hubble expansion is minimized. An arbitrary local boost (2) of the central observer leads to inferred redshifts, z_i', in the new frame given by

$$cz_i \rightarrow cz_i' = c(\gamma - 1) + \gamma \left[cz_i + v \cos\phi_i (1 + z_i) \right]$$

$$\simeq cz_i + v\cos\phi_i(1 + z_i) + \mathrm{O}\left(\frac{v^2}{c^2}\right) \tag{10}$$

$$\simeq cz_i + v\cos\phi_i + \mathrm{O}\left(\frac{v}{c}\right) , \tag{11}$$

where $\gamma = \left[1 - v^2/c^2\right]^{-1/2}$. We will adopt the widely used Newtonian velocity addition approximation (11), which is also assumed in (1). This ignores the $\mathrm{O}(v/c)$ correction, which is at most 0.5% for the boosts considered here, at least one order of magnitude smaller than typical distance uncertainties. This results in the changes $(cz_i)^2 \rightarrow (cz_i')^2 = (cz_i)^2 + 2cz_i v\cos\phi_i + v^2\cos^2\phi_i$ in the numerator of (4), and $cz_i r_i \rightarrow cz_i r_i + r_i v\cos\phi_i$ in the denominator. The linear contributions to the transformed quantities in the denominator and numerator of (4) should be approximately self–cancelling for spherical averages of data which is uniformly distributed on the celestial sphere in each shell: on average each positive contribution from the term $v\cos\phi_i$ will cancel with a negative contribution from a data point on the opposite side of the sky[4]. With such a cancellation assumed we are left with the difference

$$\Delta H_s = H_s' - H_s \sim \left(\sum_{i=1}^{N_s} \frac{(v\cos\phi_i)^2}{\sigma_i^2}\right) \left(\sum_{i=1}^{N_s} \frac{cz_i r_i}{\sigma_i^2}\right)^{-1}$$

$$\approx \frac{v^2}{2\bar{H}_0 \langle r_i^2 \rangle_s} \tag{12}$$

where

$$\langle r_i^2 \rangle \equiv \left(\sum_{i=1}^{N_s} \frac{r_i^2}{\sigma_i^2}\right) \left(\sum_{i=1}^{N_s} \frac{1}{\sigma_i^2}\right)^{-1} \tag{13}$$

is a weighted average of the squared luminosity distance in each shell. The second line of (12) follows by assuming that: (i) $\langle(v\cos\phi_i)^2\rangle = v^2 \langle\phi_i^2\rangle \sim \frac{1}{2}v^2$ is roughly constant from shell to shell; (ii) the leading order linear Hubble approximation $cz_i \simeq \bar{H}_0 r_i$ is made in the denominator. The uncertainty in (13) is given by

$$\bar{\sigma}_{\langle r_i^2 \rangle_s} = 2 \left(\sum_{i=1}^{N_s} \frac{r_i^2}{\sigma_i^2}\right)^{1/2} \left(\sum_{i=1}^{N_s} \frac{1}{\sigma_i^2}\right)^{-1} . \tag{14}$$

Our first goal is to verify that the difference in the spherically averaged Hubble expansion between the LG and CMB frames of reference is statistically consistent with a systematic variation of the form (12). To achieve this we fit a power

---

[4] The lack of data in the Zone of Avoidance does not pose a problem as this absence of data is symmetrical on opposite sides of the sky. A rough self–cancellation of the linear contributions would only be invalid when one side of the sky has a significant lack of data as compared to the opposite side of the sky. The COMPOSITE sample does indeed have sufficient sky coverage to satisfy this requirement, with the exception of the first unprimed shell 1, with $2 < r \leqslant 12.5\,h^{-1}\mathrm{Mpc}$ (Wiltshire et al. 2013), which is excluded in the data analysis.

law to the observed data. As there is a correlated uncertainty in both the independent and dependent variables, $\langle r_i^2 \rangle_s$ and ΔH_s respectively, a standard least squares method is not appropriate. Instead we use a total least squares fit, or "error in variables" method with a model of the form

$$\Delta H_s = A \left(\langle r_i^2 \rangle_s\right)^p . \tag{15}$$

In comparison with (12), we expect $p \approx -1$ and $A \approx v^2/(2\bar{H}_0)$. The details of this method are presented in the Appendix.

Carrying out this analysis, we do indeed find a difference consistent with a systematic boost offset between the LG and CMB frames of reference. Systematic uncertainties arise in the choice of shell boundaries. Considering only primed shells gives a value of $p = -1.01 \pm 0.27$. If we take the unprimed shells then we obtain $p = -0.79 \pm 0.16$ if shell 1, with $2 < r \leqslant 12.5\,h^{-1}\mathrm{Mpc}$, is included and $p = -0.77 \pm 0.35$ if this first shell – which may have insufficient sky coverage (Wiltshire et al. 2013) – is excluded. The data in the range $6.25 < r \leqslant 12.5\,h^{-1}\mathrm{Mpc}$ common to both the first primed and unprimed shells is important in establishing the boost offset which is more pronounced at small r. To account for systematic uncertainties, we have therefore applied a continuous variation of the first shell boundary in the range $2 - 6.35\,h^{-1}\mathrm{Mpc}$, while keeping the widths of the shells fixed. This leads to a value of $p = -0.88 \pm (0.25)_{\mathrm{stat}} \pm (0.13)_{\mathrm{sys}}$ where the first uncertainty is the statistical and the second systematic. For the case of the primed shells we also note the corresponding velocity calculated from the best fit value of A is $v = 772^{+1024}_{-440}$ km sec^{-1}, which is indeed close to the actual boost magnitude of 635 ± 38 km sec^{-1}, albeit with a very large uncertainty.

We repeated the analysis using 8 shells rather than 11 to smooth out variations that could interfere with the systematic boost offset. The second configuration uses shells of width $18.75\,h^{-1}\mathrm{Mpc}$, starting from an inner cutoff of $2\,h^{-1}\mathrm{Mpc}$ and $9.375\,h^{-1}\mathrm{Mpc}$ for unprimed and primed shells respectively. We find $p = -0.89 \pm 0.34$ for the primed shells and $p = -0.96 \pm 0.26$ for the unprimed shells. With a continuous variation of the inner shell boundary from $2 - 9.375\,h^{-1}\mathrm{Mpc}$, we arrive at a value of $p = -0.87 \pm (0.33)_{\mathrm{stat}} \pm (0.09)_{\mathrm{sys}}$. Fig. 2(b) shows the resultant best fit curves.

In Fig. 2 we note a discrepancy between the best fit power law and the negative values on shells in the range of $40 - 60\,h^{-1}\mathrm{Mpc}$ (or $\langle r_i^2 \rangle = 1600-3600(\,h^{-1}\mathrm{Mpc})^2$). This is understood to be the result of structures in this particular range, which give rise to both a residual monopole and dipole variation of the Hubble expansion in the LG frame. As Wiltshire et al. (2013) show, there is evidence that the boost to the CMB frame somewhat compensates for structures in this range. One finds that in the range $40 \lesssim r \lesssim 60\,h^{-1}\mathrm{Mpc}$ (and only in this range) the monopole variation is less in the CMB frame, $(\delta H_s)_{\mathrm{CMB}} < (\delta H_s)_{\mathrm{LG}}$, while the dipole magnitude is also less in the CMB frame, becoming consistent with zero in the middle of the range. If the boost to the CMB frame exactly compensated for structures in the range $40 \lesssim r \lesssim 60\,h^{-1}\mathrm{Mpc}$ then the dipole magnitude should remain close to zero in shells at larger distances. However, the magnitude of the CMB frame dipole increases to a residual offset at large distances, while the LG frame dipole is consistent with zero in most outer shells.

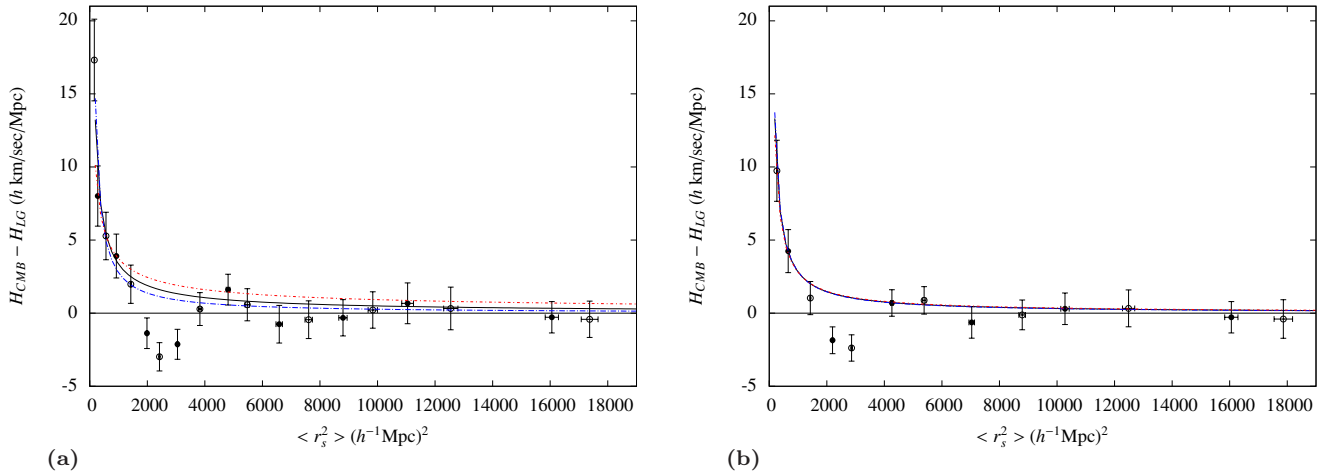Thus it appears that the dipole almost – but not en-

Figure 2. Best fit power law to the radial variation in the spherically averaged Hubble law in 2 configurations of: **(a)** 11 shells; **(b)** 8 shells. The dashed blue (lower) curve indicates the best fit to primed shells only (empty circles), the dotted red (upper) curve indicates the best fit to unprimed shells only (filled circles). The solid black curve is the combined best fit using both primed and unprimed shells. The first data point – corresponding to unprimed shell 1 – is omitted in each plot, as it is off the scale.

tirely – has the character of a Lorentz boost dipole. Given that there are structures that give rise to a residual nonlinear Hubble expansion in the range $40 \lesssim r \lesssim 60\,h^{-1}\mathrm{Mpc}$ we cannot expect a perfect power law fit (15) in Fig. 2. However, the deviation from a power law is consistent with the observation that $(H_s)_{\mathrm{CMB}} < (H_s)_{\mathrm{LG}}$ in the range over which the boost almost compensates for nonlinear structures.

Now that we have verified the power law nature of the difference $H_{\mathrm{CMB}} - H_{\mathrm{LG}}$ we must check whether this result is unique for the boost to the LG frame. To investigate this we determine the Hubble constant in radial shells for frames boosted arbitrarily with respect to the CMB, denoted by frame "X", and then fit (15) to the resulting $\Delta H = H_{\mathrm{CMB}} - H_{\mathrm{X}}$ curve. We vary the direction of the boost to frame X while holding the magnitude constant, thus producing a sky map. We first choose a magnitude of $635\,\mathrm{km\ sec^{-1}}$ corresponding to the boost from the CMB to LG frame of reference.

To display these sky maps in a meaningful fashion we cannot simply plot the value of p. Suppose that the CMB is boosted from a frame which has $\Delta H_s = A(\langle r_i^2 \rangle_s)^p$ with $p = -1$ and $A > 0$, representing the best fit boost offset. If one now boosts in the *opposite* direction by $635\,\mathrm{km\ sec^{-1}}$ then one finds a best fit power law with $\Delta H_s = A(\langle r_i^2 \rangle_s)^p$ with $p \approx -1$ but $A < 0$ since the CMB frame necessarily has the smaller value of H_s on average. In each case we must first of all determine whether (15) gives a better overall fit with $A > 0$ or $A < 0$ – given that *some* data points will always be opposite to the overall trend. In Fig. 3 we plot[5]

$$f_p = \begin{cases} |p+1|, & A \geqslant 0 \\ 2 - |p+1|, & A < 0 \end{cases} \qquad (16)$$

which takes the value $f_p = 0$ at the best fit with $A > 0$ and $f_p = 2$ at the best fit with $A < 0$. The latter point turns out

to be in the opposite direction, but not exactly opposite the best fit direction, reflecting the uncertainties in the method.

Fig. 3 shows that for a boost magnitude of $635\,\mathrm{km\ sec^{-1}}$ the Local Group is indeed contained in a set of frames that display strong evidence of being a minimum Hubble expansion variation frame as $p \approx -1$. More precisely, the monopole Hubble expansion in the CMB frame compared to frames boosted at $635\,\mathrm{km\ sec^{-1}}$ relative to the CMB has the mathematical character of a systematic boost offset for directions close to that of the LG. In Fig. 3 we can see a distinctive difference between the directions for which the boosted frame has the lesser variation ($A > 0$, values plotted closest to 0), and the directions for which the CMB frame has the lesser variation ($A < 0$, values plotted closest to 2).

We have verified that the additional monopole variation of the Hubble expansion seen in the CMB frame does have the character of what is expected by a boost from the LG frame, if the LG frame is close to the frame in which the monopole variation is minimized. However, as yet we have not varied the magnitude of the boost. Given the uncertainties we have already noted, it is clear that any systematic boost offset will become very hard to confirm statistically if the boost magnitude is small, since the other uncertainties will then become dominant. The systematic boost offset method can only give a rough indication of the boost direction for the large boosts rather than defining a precise "minimum Hubble expansion variation frame". There are in fact many degenerate frames.

4 MINIMIZING THE AVERAGE SPHERICAL HUBBLE EXPANSION VARIATION

We will now investigate arbitrary variations of the boost magnitude and direction to determine to what extent we can define a frame of reference in which the monopole variation in the Hubble expansion is minimized.

---

[5] Both primed and unprimed shells are used (in the 11 shell case), to produce a smoothed sky map without determining systematic uncertainties.
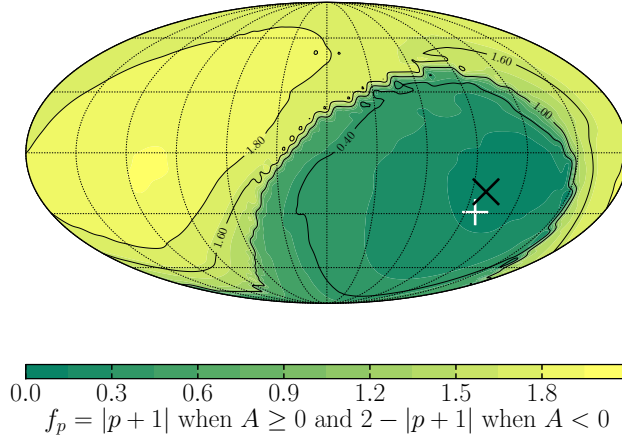
Figure 3. The best fit parameters to a systematic boost offset (15) for frames boosted from the CMB frame at $635\,\mathrm{km\ sec^{-1}}$. The black cross denotes the boost to the LG frame and the white cross denotes the boost to the frame with minimum variation in the spherically averaged Hubble law for a boost of this magnitude, which will be discussed in §4.1. In all figures, the galactic longitudes $\ell = 0°, 180°, 360°$ are on the right edge, centre and left edge respectively.

4.1 Variations in the "nonlinear regime"

Initially we will consider monopole variations with respect to a uniform $\delta H = 0$ expectation below the scale of statistical homogeneity ($\lesssim 100\,h^{-1}\mathrm{Mpc}$). This is quantified by summing the mean square differences of (9), to give a statistic

$$\chi_{\mathrm{a}}^2(n_f, n_i) = \sum_{i=n_i}^{n_f} \frac{\bar{H}_0^4 \delta H_i^2}{\bar{H}_0^2 \sigma_{H_i}^2 + H_i^2 \sigma_{\bar{H}_0}^2}, \qquad (17)$$

where n_f and n_i define the upper and lower shells included in the range of the calculation respectively. We will take the primed shells with, $n_i = 1'$ and $n_f = 8'$, covering the range $6.25 < r \leqslant 106.25\,h^{-1}\mathrm{Mpc}$. This includes all data potentially in the regime of nonlinear Hubble expansion, while excluding the innermost unprimed shell which may have incomplete sky coverage.

In identifying a frame of "minimum nonlinear Hubble variation" we must also estimate an uncertainty about this minimum. Since we do not expect an exact fit to a linear Hubble expansion in *any* frame it is inappropriate to use the ordinary confidence intervals of the χ_{a}^2 distribution, as this would not give a fair measure of the improvement in uniformity between two reference frames, instead rendering all reference frames at least[6] $1\,\sigma$ from the "expectation". Instead we follow Wiltshire et al. (2013) in using Bayesian statistics to give a measure of the relative probability of uniformity of Hubble expansion in different reference frames. In particular, using the complementary incomplete gamma function for the χ_{a}^2 distribution we directly calculate the probabilities P_{U}, P_{V} of the spherically averaged Hubble law giving a value that coincides with $\bar{H}_0$, using the shells $1'$ to $8'$, for the pair of reference frames U and V. A Bayes factor $B = P_{\mathrm{U}}/P_{\mathrm{V}}$ is then computed. We use the bounds $\ln B = 1, 3, 5$ to denote

confidence regions, where the range $1 < \ln B \leqslant 3$ represents *positive* Bayesian evidence, $3 < \ln B \leqslant 5$ represents *strong* Bayesian evidence and $\ln B > 5$ represents *very strong* evidence (Kass & Raftery 1995). Thus if we take frame U to be the frame with the minimum χ_{a}^2 within a specified set of reference frames, any boosted frame V with Bayes factor $\ln B \leqslant 1$ relative to U can be considered statistically equivalent to U in the sense the difference in probability is *"not worth more than a bare mention"* (Kass & Raftery 1995).

The frame of reference with the minimum monopole variation is found using a downhill optimization with (17). This reveals a global minimum variation (MV) frame for a boost in the direction $(\ell, b) = (59.3°, 16.6°)$ with magnitude $740.6^{+258.4}_{-487.0}\,\mathrm{km\ sec^{-1}}$ with respect to the LG frame. (The uncertainty is taken from the $\ln B \leqslant 1$ bound.) Even though the boost is large, the uncertainties are also large. The Bayes factor $\ln\left(P_{\mathrm{MV}}/P_{\mathrm{LG}}\right) = 2.77$ constitutes positive but not strong evidence for the MV frame relative to the LG frame. However, the corresponding boost from the CMB reference frame to the global minimum frame has a magnitude $1203^{+375}_{-458}\,\mathrm{km\ sec^{-1}}$, which is even larger and now very significantly nonzero, with $\ln\left(P_{\mathrm{MV}}/P_{\mathrm{CMB}}\right) = 33.3$, which is extremely strong evidence against the CMB frame.

To visualize the distribution of χ_{a}^2 in the 3-dimensional parameter space $\{v, \ell, b\}$ we show two angular slices at fixed values of v in Fig. 4, and a slice along the locus of (ℓ, b) values for which χ_{a}^2 is minimized for fixed v in Fig. 5. The angular distribution in Fig. 4 is similar for both velocities. However, the Bayesian confidence regions grow large as v is decreased, eventually taking up the whole sky for very small velocities. Only for large boosts is there a well-defined boost direction that reduces monopole variation.

In Fig. 5 for each boost magnitude we locate the direction of minimum χ_{a}^2 and plot the corresponding value. The angular coordinates (ℓ, b) of the minimum are different in each case. The distribution of χ_{a}^2 relative to the $\ln B \leqslant 1$ confidence interval (dashed vertical lines on Fig. 5(a)) re-

---

[6] At the global minimum we find a χ_{a}^2 per degree of freedom of 2.0.
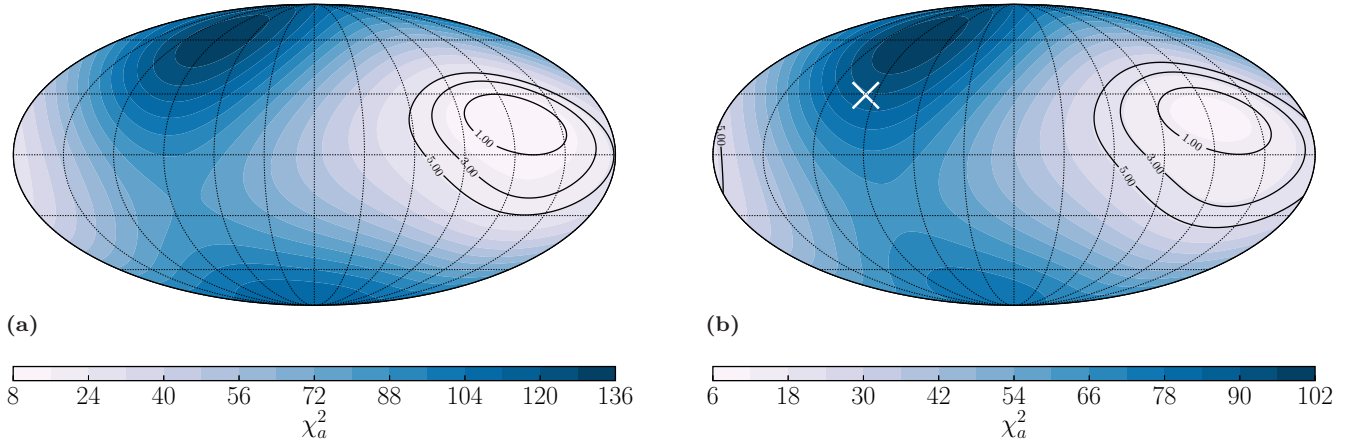
Figure 4. Contour maps of angular variation of χ^2_a for two choices of boost magnitude with respect to the Local Group: **(a)** 740 km sec^{-1} **(b)** 635 km sec^{-1}. The solid contours show the levels of Bayesian evidence $\ln B = \ln(P_{\min}/P_{(\ell,b)})$ for each direction (ℓ, b) with respect to the frame of minimum χ^2_a in each case. The white cross on (b) shows the direction of the boost to the CMB frame (also of magnitude ≈ 635 km sec^{-1}). In all figures, the galactic longitudes $\ell = 0°, 180°, 360°$ are on the right edge, centre and left edge respectively.
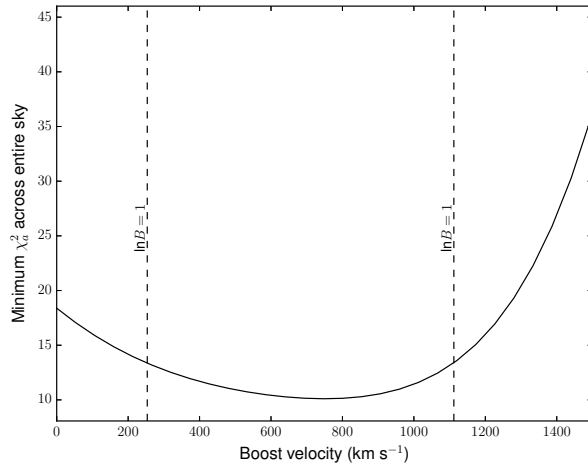

Figure 5. Variation of the minimum χ^2_a with fixed boost velocity, v. The locus of (ℓ, b) values for which this minimum is found lies within $\approx 20°$ of the galactic plane for cases in which $\ln B = \ln(P_{\mathrm{MV}}/P_{\mathrm{V}}) \leqslant 1$ as indicated by the dashed line (where MV is the global minimum and V any frame along the locus).

veals the main problem in constraining a "minimum variation frame" by this technique. The near flat distribution of χ^2_a values within the $\ln B \leqslant 1$ bound of the global minimum are found to lie on a locus of boost directions (ℓ, b) which all lie close to the galactic plane. This is of course the Zone of Avoidance region, where the COMPOSITE sample lacks data, due to the Milky Way obscuring distant galaxies. Evidently, we are free to perform large boosts in the plane of the galaxy, as the data is not constrained there. This hypothesis could be checked by simulating data with the same characteristics as the COMPOSITE sample, using exact solutions of Einstein's equations (Bolejko et al. 2015). Such an investigation is beyond the scope of this paper. For now, with

the available data we can only conclude that the LG frame is not ruled out as the standard of rest by this criterion, since the Bayesian distinction between the two frames with $|\ln B| = 2.77$ is not strong.

We have found a set of degenerate frames of reference which might be taken as the minimum average monopole variation frame. Although large uncertainties still exist, we are able to see that the boost to the CMB frame, shown by a white cross in Fig. 4(a), is far from our degenerate set of possible boosts to the minimum variation frame. In Table 1 we make a comparison similar to Table 1 of Wiltshire et al. (2013) between the LG frame and the global minimum χ^2_a (MV) frame. In particular, we directly calculate the prob-

Table 1. Hubble expansion variation in radial shells in minimum Hubble expansion variation (MV) and LG frames. Spherical averages (4) are computed for two different choices of shells, $r_s < r \leqslant r_{s+1}$, the second choice being labeled by primes. In each case we tabulate the inner shell radius, r_s; the weighted mean distance, $\bar{r}_s$; the shell Hubble constants, $(H_s)_{\mathrm{LG}}$ and $(H_s)_{\mathrm{MV}}$ in the LG and MV frames, and their uncertainties determined by linear regression within each shell, together with its "goodness of fit" probability Q_s and reduced χ^2 (for $\nu_s = N_s - 1$); $\ln B$ where B is the Bayes factor for the relative probability that the MV frame has more uniform $\delta H_s = 0$ than the LG frame when χ^2 is summed in all shells with $r > r_s$. H_s and $\bar{\sigma}_s$ are given in units $h\,\mathrm{km\ sec^{-1}\,Mpc^{-1}}$.

Shell s	1	2	3	4	5	6	7	8	9	10	11
N_s	92	505	514	731	819	562	414	304	222	280	91
r_s (h^{-1}Mpc)	2.00	12.50	25.00	37.50	50.00	62.50	75.00	87.50	100.00	112.50	156.25
$\bar{r}_s$ (h^{-1}Mpc)	5.43	16.33	30.18	44.48	55.12	69.24	81.06	93.75	105.04	126.27	182.59
$(H_s)_{LG}$	117.9	103.1	106.5	105.5	104.8	102.1	102.8	103.2	103.7	102.4	101.0
$(\bar{\sigma}_s)_{LG}$	4.6	1.4	1.0	0.7	0.7	0.7	0.9	0.9	1.0	0.8	1.7
$(Q_s)_{LG}$	0.000	0.000	0.000	0.000	0.998	0.940	1.000	1.000	1.000	0.993	0.999
$(\chi_s^2/\nu_s)_{LG}$	23.656	7.767	2.185	1.419	0.864	0.909	0.594	0.542	0.622	0.803	0.590
$(H_s)_{MV}$	118.5	102.9	106.7	104.5	104.8	102.9	102.6	103.9	104.9	102.7	102.0
$(\bar{\sigma}_s)_{MV}$	4.6	1.4	1.1	0.7	0.7	0.7	0.9	0.9	1.0	0.8	1.7
$(Q_s)_{MV}$	0.000	0.000	0.000	0.000	0.330	0.887	1.000	1.000	1.000	0.964	0.999
$(\chi_s^2/\nu_s)_{MV}$	29.130	12.320	3.037	2.005	1.021	0.928	0.682	0.600	0.667	0.854	0.603
$\ln B$ ($r \geqslant r_s$)	3.53	2.85	2.79	1.99	0.85	0.33	0.36	0.14	0.07	0.45	

Shell s	$1'$	$2'$	$3'$	$4'$	$5'$	$6'$	$7'$	$8'$	$9'$	$10'$	11
N_s	321	513	553	893	681	485	343	273	164	206	91
r_s (h^{-1}Mpc)	6.25	18.75	31.25	43.75	56.25	68.75	81.25	93.75	106.25	118.75	156.25
$\bar{r}_s$ (h^{-1}Mpc)	12.26	23.46	37.61	49.11	61.74	73.92	87.15	99.12	111.95	131.49	182.59
$(H_s)_{LG}$	103.5	103.5	103.9	106.6	103.9	102.0	103.2	103.6	101.6	102.7	101.0
$(\bar{\sigma}_s)_{LG}$	1.8	1.1	0.9	0.7	0.8	0.8	0.9	0.9	1.0	0.9	1.7
$(Q_s)_{LG}$	0.000	0.000	0.000	0.031	0.960	1.000	1.000	1.000	0.996	0.999	0.999
$(\chi_s^2/\nu_s)_{LG}$	11.427	3.246	1.792	1.090	0.907	0.701	0.592	0.608	0.728	0.711	0.590
$(H_s)_{MV}$	102.7	104.3	103.3	106.1	104.2	102.9	102.9	104.7	102.7	102.9	102.0
$(\bar{\sigma}_s)_{MV}$	1.8	1.1	0.9	0.7	0.8	0.8	0.9	0.9	1.0	0.9	1.7
$(Q_s)_{MV}$	0.000	0.000	0.000	0.000	0.481	1.000	1.000	1.000	0.967	0.997	0.999
$(\chi_s^2/\nu_s)_{MV}$	18.547	4.940	2.429	1.428	1.002	0.734	0.704	0.613	0.807	0.752	0.603
$\ln B$ ($r \geqslant r_s$)	2.38	2.32	2.34	1.93	0.55	0.33	0.41	0.14	0.24	0.50	

abilities P_{LG}, P_{MV} of the spherically averaged Hubble law giving a value that coincides with $\bar{H}_0$, when all shells larger than a given shell are included. If we exclude the innermost unprimed shell 1, then the Bayesian evidence for the MV frame having a more uniform average Hubble expansion than the LG frame, lies at most in the range $1 < \ln B \leqslant 3$, which represents positive but not strong Bayesian evidence (Kass & Raftery 1995). By contrast, the Bayesian evidence that the LG frame has a more uniform spherically average Hubble expansion than the CMB frame is very strong with $\ln B > 5$ (Wiltshire et al. 2013).

4.2 Variations in the "linear regime"

The goodness of fit, Q_s, of a linear Hubble law in the innermost shells of Table 1 is poor, as we would expect since these shells are in the *nonlinear* regime. Beyond approximately $75\,h^{-1}$Mpc we expect to pass into the linear regime (Scrimgeour et al. 2012), and this is seen with the decreasing values of χ_s^2/ν_s, where χ_s^2 is given by (3) for each shell.

However, contrary to expectation, the MV frame determined from (17) has an overall poorer goodness of fit Q_s than the LG frame to a linear Hubble law in shells with $s \geqslant 4'$, including in particular in shells $9'$ and 10 which should be in the linear regime. We note that the asymptotic value, $\bar{H}_0$, is 1% higher in the MV frame as compared to the LG frame, and this may contribute to δH_s being smaller, even though the goodness of fit in individual shells is poorer in some cases[7].

Any true candidate for a minimum Hubble expansion variation frame should also clearly demonstrate the emergence of a linear Hubble law consistent with the existence of a statistical homogeneity scale. The χ_a^2 statistic (17) involves minimizing the variation $\delta H_s = (H_s - \bar{H}_0)/\bar{H}_0$ relative to the asymptotic Hubble constant. But a boost can also alter

---

[7] The value of $\bar{H}_0$ is also 1% larger in the LG frame as compared to the CMB frame, but by contrast the goodness of fit in the linear regime shells is better in the LG frame than in the CMB frame (Wiltshire et al. 2013).
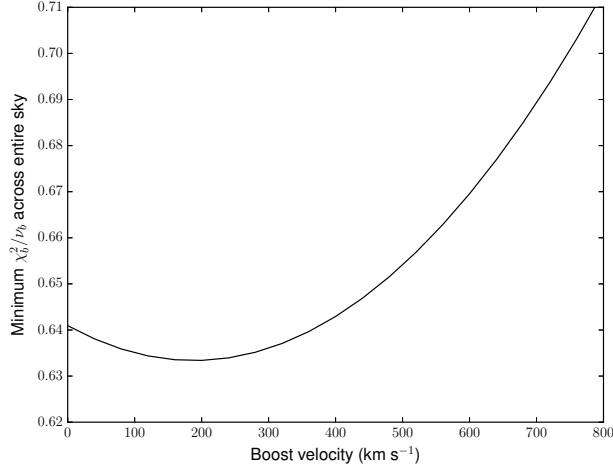
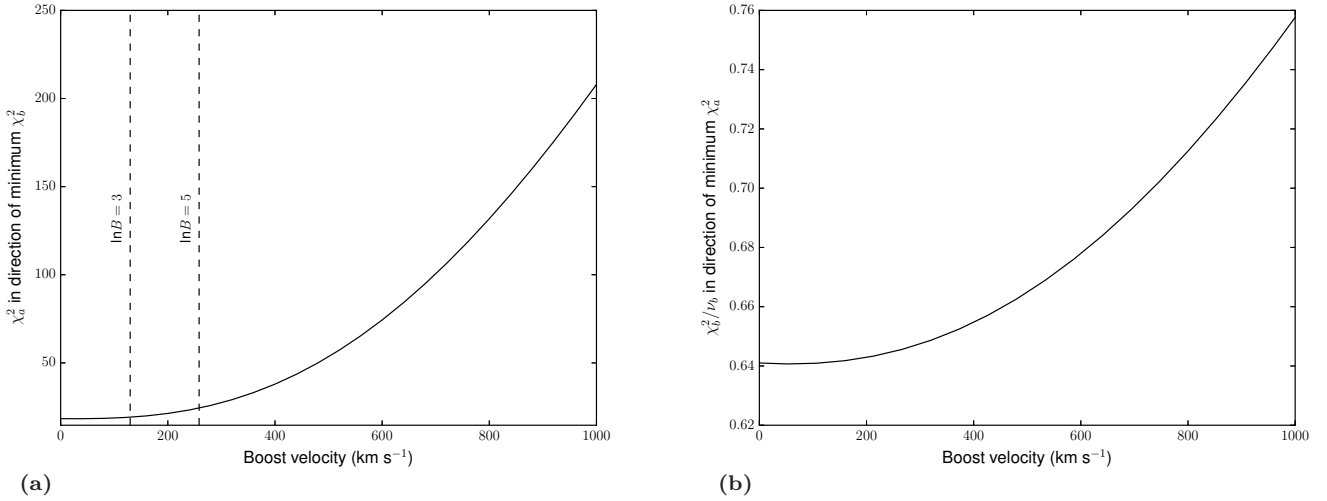**Figure 6.** Variation of the minimum χ_b^2/ν_b with fixed boost velocity, v.

(a)

(b)

Figure 7. (a) Variation of χ_a^2 along a locus of (ℓ, b) values for which χ_b^2 is minimized at each fixed boost velocity v. **(b)** Variation of χ_b^2/ν_b along a locus of (ℓ, b) values for which χ_a^2 is minimized at each fixed boost velocity v.

$\bar{H}_0$ in a way which makes for a worse goodness of fit to a linear Hubble law. Therefore it is not a completely suitable candidate statistic.

In order to quantify the emergence of a linear Hubble law we will therefore alternatively minimize the quantity

$$\chi_b^2 = \sum_{s=7}^{11} \chi_s^2 \qquad (18)$$

where χ_s^2 is given by (3) in the sth shell. This sum is performed over the unprimed configuration of shells as shell 7 has an inner cutoff near the boundary of the nonlinear and linear regimes. Thus (18) gives a measure of the goodness of fit to a linear Hubble law averaged over the outer 5 shells, without normalizing the asymptotic Hubble constant. Defining the total degrees of freedom $\nu_b = \left(\sum_{s=7}^{11} N_s\right) - 1$, we find $\chi_b^2/\nu_b = 0.631$ for the LG frame, $\chi_b^2/\nu_b = 0.692$ for the MV (minimum χ_a^2) frame, and $\chi_b^2/\nu_b = 0.653$ for the CMB frame. Thus even the CMB frame shows a clearer emerging linear Hubble law than the minimum χ_a^2 frame.

To determine whether there is any frame with a more definitive emerging linear Hubble law than the LG frame, we determine the distribution of (18) upon making arbitrary boosts with respect to the LG frame.

In Fig. 6 we locate the direction of minimum χ_b^2 at each boost magnitude and plot the corresponding value, analogously to Fig. 5. We find a best fit boost of 222.3 km sec^{-1} in the direction $(\ell, b) = (241.84°, 70.53°)$ with respect to the LG frame, with a value of $\chi_b^2/\nu_b = 0.621$. This is 45° from the direction of the residual CMB temperature dipole in the LG frame, and so does not appear related.

In Fig. 7 for each boost magnitude we calculate the value of χ_a^2 in an (ℓ, b) direction determined by minimizing with respect to χ_b^2, and vice versa. Thus, we compute the locus of (ℓ, b) values in the $\{v, \ell, b\}$ parameter space that minimize χ_b^2 for each fixed v, and then compute χ_a^2 at these parameter values, and vice versa. It is apparent that making boosts of the order $\gtrsim 100$ km sec^{-1} along the locus of (ℓ, b) values which minimize χ_b^2 results in an increase in χ_a^2 to

values which are disfavoured by Bayesian evidence. This is because making improvements in χ_b^2 requires boosts in (ℓ, b) directions away from our minimum χ_a^2.

However, we note that: (i) all values of χ_b^2 shown in Fig. 6 and 7 are consistent with very probable fits, given the low values of χ_b^2/ν_b; (ii) there are less data points per se in those outer shells used in χ_b^2 as compared to the inner shells used in χ_a^2; and (iii) local boosts have a relatively small effect on the values of cz_i in the outer shell. These facts mean we should also exercise caution about drawing strong conclusion solely from the minimization of χ_b^2. The primary utility of the χ_b^2 statistic is not as a sole discriminator, but as a check against the potential bias in the χ_a^2 statistic due to its anchoring to the value $\bar{H}_0$ in the outermost shell.

Finding a joint minimum for χ_a^2 and χ_b^2 would be feasible if the (ℓ, b) values of each global minimum were close on the sky. However, this is not the case – the global minimum for χ_a^2 is at an angle of $93°$ from the global minimum of χ_b^2, i.e., they are roughly orthogonal. As there is no unbiased way to weight these two statistics we cannot set out to determine a weighted minimum as any result would be highly sensitive to our choice of weighting.

4.3   Systematic boost offsets from the Local Group

Neither statistic χ_a^2 nor χ_b^2 appears entirely satisfactory for establishing a global minimum expansion variation frame. The χ_a^2 statistic is the better measure of Hubble expansion variation in the nonlinear regime but is also affected by potential bias in the anchoring of $\bar{H}_0$. The most we can say is that there is a freedom to perform large boosts in the plane of the galaxy, given the lack of data in the Zone of Avoidance.

If χ_a^2 is taken as the better statistic, then a criterion for breaking the boost degeneracy may be possible by returning to the systematic boost offset analysis of §3. Any true best fit frame should show a clear signal of a boost offset (12) with respect to the Local Group. The "best" boost offset can be characterized in 3 ways, each with its own challenges:

(1) Determine the boost for which $p = -1$. This is hindered by the fact that there are many boosts that satisfy this criterion, at almost every magnitude from the LG.
(2) From the value of A in (15) determine a derived boost velocity, v_{der}. Any boost offset should have v_{der} consistent with the true boost magnitude v_{true} within uncertainties. However, this is difficult due to the large uncertainties associated with the value of A.
(3) Determine a measure of variation in the fit of the boost offset, given by (A8) in the Appendix. This is also problematic since all fits are extremely good due to the large uncertainties in the H_s.

We will therefore use method (1) to determine the direction of the boost on the sky, and then consider (2) and (3) to constrain the magnitude[8].

---

[8] The choice of shell boundaries introduces systematic uncertainties in this analysis. In the production of the sky maps in Fig. 8 we calculate both primed and unprimed shells and fit the power law (15) to all points. For Fig. 9 we consider fits to primed and unprimed only, and to both.

First we check for a systematic boost offset for the global χ_a^2 minimum frame determined in §4.1. A sky map of f_p values as given by (16) for boosts of magnitude $740 \, \mathrm{km \, sec^{-1}}$ is given in Fig. 8(a), with the $\ln B \leqslant 1$ and $1 < \ln B \leqslant 3$ confidence regions for χ_a^2 displayed. We note that there are in fact boosts with values of $p \approx -1$ consistent with the χ_a^2 minimum. However, these directions are far more constrained and do not align with the exact minimum. In addition, the (ℓ, b) direction at this magnitude with the best fit to (15) has an inconsistent value of the derived velocity. Thus we do not see a clear systematic boost offset between the Local Group and the frame corresponding to the global minimum χ_a^2, further ruling this out as a potential candidate for the standard of rest we are looking for.

The next step is a global search for the best systematic boost offset from the LG. In order to further understand the angular distribution of f_p values (16) for boosts from the LG frame, we arbitrarily choose a boost magnitude of $200 \, \mathrm{km \, sec^{-1}}$ and plot f_p with respect to (ℓ, b) in Fig. 8(b). We have found that for all interpolating velocities between the $200 \, \mathrm{km \, sec^{-1}}$ and $740 \, \mathrm{km \, sec^{-1}}$ cases displayed in Figure 8 there is a region of (ℓ, b) values for which $f_p \approx 0$. Thus in order to use this method to find a realistic systematic boost offset we must use an additional criterion.

In Fig. 9 we plot the values of v_{der}/v_{true} and $S/(n-2)$ (for n data points) from (A8) with respect to the boost magnitude, where for each magnitude the (ℓ, b) direction is that for which p is closest to -1 and $A > 0$ (i.e. $f_p \approx 0$). Thus, we can use these additional quantities to constrain a systematic boost offset along the locus of (ℓ, b) directions in the 3-dimensional $\{v, \ell, b\}$ parameter space. The expected value of S has a χ^2 distribution for $(n-2)$ degrees of freedom, and thus $S/(n-2)$ has an expectation value of unity (York et al. 2004). Clearly, the values of $S/(n-2)$ in Fig. 9(b) are consistent with a very good fit to (15) for all boosts. Our inability to tightly constrain the boost magnitude is no doubt due to the lack of data in the Zone of Avoidance and large uncertainties in the values of $H_{LG} - H_X$. Although (A8) it is not useful for constraining the boost magnitude, we nonetheless see that the ratio of derived and true velocities in Fig. 9 does show a meaningful difference on this interval.

Using both primed and unprimed shells in our calculation we find that the $v_{der}/v_{true} = 1$ at $v_{true} = 171 \, \mathrm{km \, sec^{-1}}$ in a direction $(\ell, b) = (59°, -4°) \pm (5°, 4°)$. For this boost, we find the result $p = -1.18 \pm (0.95)_{stat} \pm (0.32)_{sys}$, where the systematic uncertainty is determined as in §3, consistent with $p = -1$. The difference between the spherically averaged Hubble expansion in this frame, which we denote by X and the LG is shown in Fig. 10. One can see that a systematic boost offset is apparent here. However, due to the large uncertainties that arise when taking differences of the H_s the result is also statistically consistent with zero. (Thus the question of whether the first unprimed shell 1 should be included in the analysis, due to its incomplete sky cover, is immaterial. Indeed, if we take only primed shells then we obtain $p = -1.17 \pm (1.2)_{stat}$.) This frame X is only weakly disfavoured compared to the global χ_a^2 minimum, $\ln(P_{MV}/P_X) = 1.7$, and is within 1σ of the global minimum of χ_b^2; similar to the LG frame on both counts.

Our choice of frame X above is based on taking $v_{der}/v_{true} = 1$, a condition which may only be approximately matched in reality, given our huge uncertainties in the co-
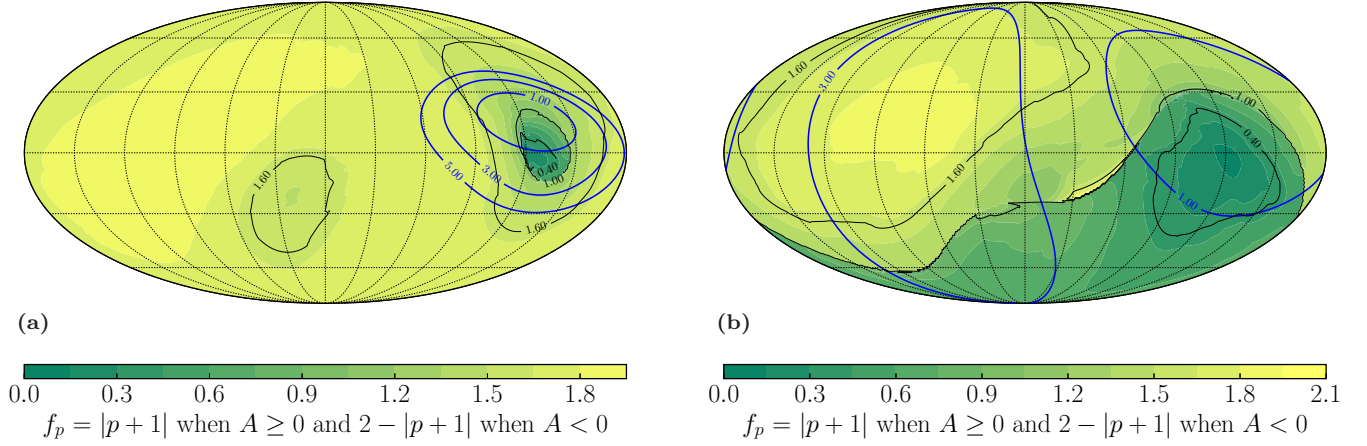
Figure 8. The best fit parameters to a systematic boost offset for boosts from the Local Group of magnitude: **(a)** 740 km sec^{-1}; **(b)** 200 km sec^{-1}. The thick blue contours denote the corresponding χ_a^2 distribution. The solid contours show the levels of Bayesian evidence $\ln B = \ln(P_{\min}/P_{(\ell,b)}) = 1,3,5$ for each direction (ℓ,b) with respect to the frame of minimum χ_a^2 at each boost magnitude. In (a) $\ln B = 1,3,5$ contours are visible while in (b) only the $\ln B = 1,3$ contours are visible, as no directions have $\ln B > 5$ in that case. In all figures, the galactic longitudes $\ell = 0°, 180°, 360°$ are on the right edge, centre and left edge respectively.
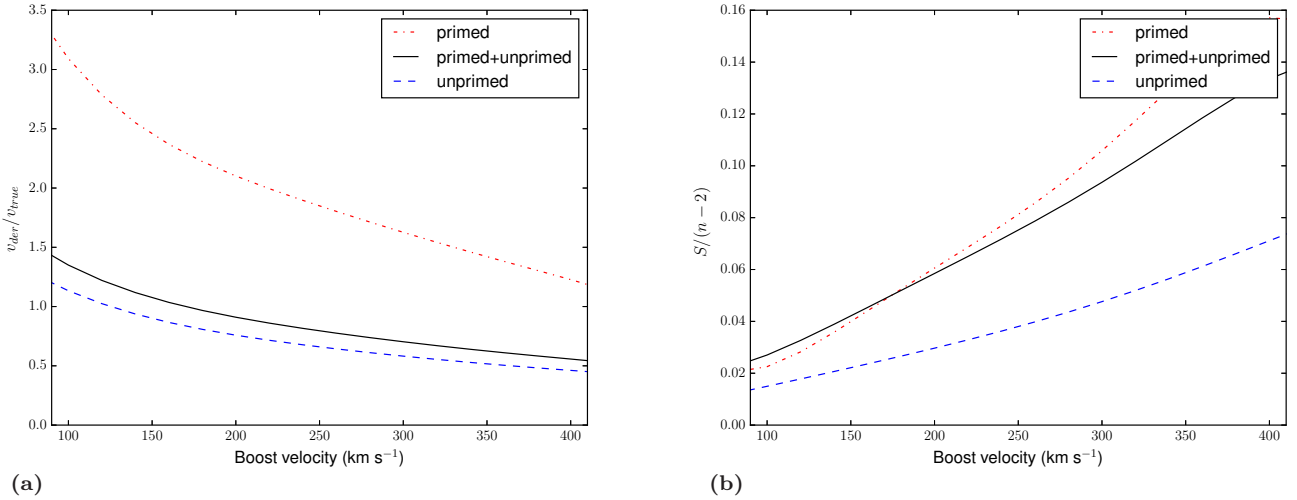

Figure 9. (a) The ratio of the derived velocity from the best fit power law and the true boost velocity, at each v the direction is determined by the best fit for which p is closest to -1. **(b)** The statistical variation in the same direction as determined by $S/(n-2)$ from (A8), where n is the number of data points being fit.

efficient A. We again have a degeneracy in the choice of minimum Hubble expansion variation frame that satisfies the two conditions $p = -1$ and $v_{\mathrm{der}} = v_{\mathrm{true}}$.

The results of this section confirm the finding of §4.1 that our determination of a suitable cosmic rest frame is limited in the COMPOSITE sample by a degeneracy under boosts close to the plane of the galaxy. The consistency between the methods of this section and §4.1 may be less significant, as they are not completely independent. In the LG frame the primary source of the monopole variation is the increased value of H_s in the innermost shells, while the

more distant shells closer to the linear regime show closer to asymptotic values. Thus boosting to a frame with a reduced H_s in the innermost shells will give the most significant improvement to χ_a^2, relative to which small changes in the more distant shells are negligible. This is precisely the type of difference we model with a power law of the form (15) with $p \approx -1$. Consequently, if our hypothesis concerning (15) is correct then it is not surprising that we see the consistency in the angular directions that minimize χ_a^2 on one hand, and which give values of $p \approx -1$ with $A > 0$ on the other.
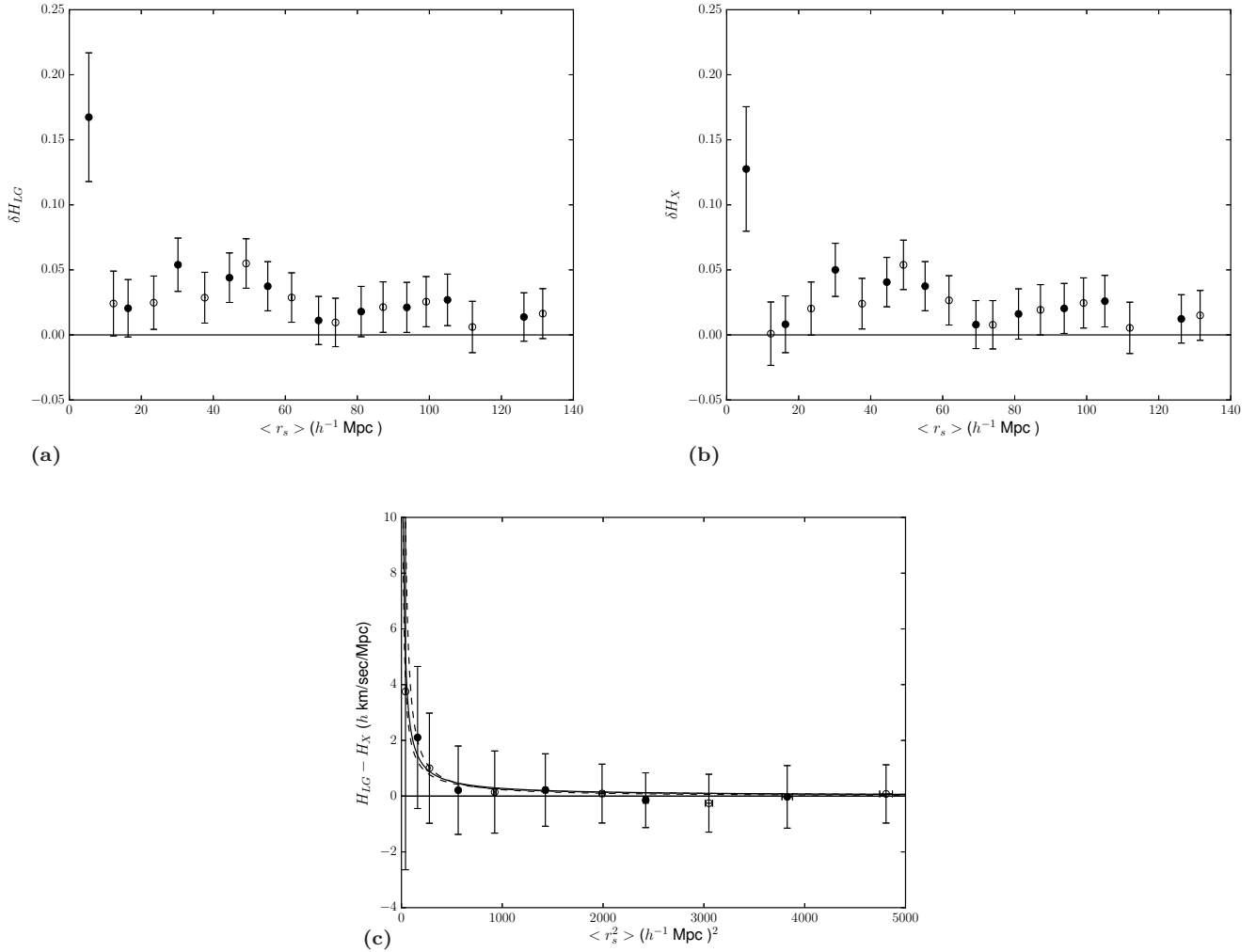
Figure 10. The variation (9) in the spherically averaged Hubble law: **(a)** δH_s in the LG frame; **(b)** δH_s in frame X; **(c)** the systematic boost offset between the LG and frame X. (Note: the first unprimed shell 1 is shown.)

4.4 Angular Hubble expansion variation

Wiltshire et al. (2013) also explore the extent to which angular averages of the Hubble expansion offer an independent characterization of a minimum Hubble variation frame of reference. When one takes angular Gaussian window averages of the Hubble expansion a dipole becomes apparent. Wiltshire et al. (2013) show that this dipole is strongly correlated with the residual CMB temperature dipole when both are referred to the LG (or LS) rest frame. If we are to define a new cosmic standard of rest, within which we still observe a residual CMB temperature dipole, then Wiltshire et al. (2013) argue that such a dipole must have a nonkinematic origin. The correlation of the residual CMB temperature dipole and Hubble expansion dipole supports the proposal that structures in the nonlinear regime of expansion are simultaneously responsible for these effects. Thus we are interested in finding a frame of reference in which this correlation is maximized.

We have investigated this question, and find that correlation of the residual CMB temperature and Hubble variation dipoles under arbitrary boosts does not offer a viable

characterization of the minimum variation frame (McKay 2015). By making boosts in the appropriate direction from the LG we are able to artificially increase both the magnitude of the Hubble expansion dipole *and* the CMB residual temperature dipole simultaneously, at the expense of also increasing the monopole expansion variation. By calculating the Hubble expansion dipole and higher multipole coefficients using HEALPIX[9] in the *frame of maximum correlation* for a given boost magnitude from the LG we can observe this artificial increase in the strength of the dipole relative to the higher multipoles. Boosting in directions which make the dipoles more pronounced will naturally increase the correlation, but this is irrelevant if the monopole variation is also increased. Therefore, as this artificially induced increase in the correlation cannot be distinguished from a physically meaningful increase, this method does not offer a viable characterization of the minimum variation frame and was abandoned as a line of investigation (McKay 2015).

---

[9] http://healpix.jpl.nasa.gov/(Gorski et al. 2005)

5 HUBBLE EXPANSION VARIATION IN THE *COSMICFLOWS*-2 CATALOGUE

Thus far our investigation has been based entirely on the COMPOSITE catalogue of distances and redshifts. In this section we aim to repeat the monopole Hubble expansion variation analysis on the recently released *Cosmicflows*-2 (CF2) catalogue. Systematic differences become apparent in this analysis which we will investigate in §6.

CF2 is a compilation of distances and redshifts from both new and existing sources of observational data. The entire catalogue consists of over 8000 galaxies both locally and extending beyond the scale of statistical homogeneity. This compilation has a large subset of galaxies and galaxy clusters in common with the COMPOSITE sample, including the large SFI++ sample of Springob et al. (2007). In total, the distances are determined by six different methods and compiled together by Tully et al. (2013). The CF2 data is presented in two sets, one with all individual galaxies included, and one condensed into galaxy groups, including groups consisting of one galaxy. We will use the entire data set of 8162 galaxy redshifts and distances, freely available from the extragalactic distance database[10].

5.1 Spherically averaged Hubble expansion and treatment of biases

We repeat our earlier analysis to calculate δH_s (9) in the same two configurations of 11 spherical shells. The CF2 data is presented with a modified "recession velocity", cz, and a raw observed redshift. Given the prevalence of the use of such modifications, particularly in bulk flow studies, it is worthwhile to briefly investigate the effect such modifications can have on monopole Hubble expansion variation. Tully et al. (2013) define an adjustment *assuming a FLRW cosmology with* $\Omega_{M0} = 0.27$ *and* $\Omega_{\Lambda0} = 0.73$. This adjustment is given by a Taylor expansion to $\mathcal{O}(z^3)$ of a homogeneous isotropic expansion law,

$$cz_{\mathrm{mod}} = cz\left[1 + \tfrac{1}{2}(1-q_0)z - \tfrac{1}{6}(1-q_0-3q_0^2+1)z^2\right], \quad (19)$$

where $q_0 = 0.5(\Omega_{M0} - 2\Omega_{\Lambda0})$ and z is the redshift in the CMB frame.

Since we wish to deal with cosmological model–independent quantities, this is not the type of adjustment that should be made. In particular, a homogeneous isotropic expansion law cannot be assumed below the scale of statistical homogeneity if the conclusions of Wiltshire et al. (2013) are correct. Nor should such an expansion law be assumed in the CMB rest frame. However, for completeness we consider the adjustment (19) in order to rule it out as the cause of much larger systematic differences we will discuss shortly.

Fig. 11 shows δH using the CF2 sample. It is immediately evident that these plots are very different to those found with the COMPOSITE sample, as given in Fig. 3 of Wiltshire et al. (2013). Most notably, the uniformity of Hubble expansion in spherical shells is considerably worse in CF2 as compared to the COMPOSITE sample, in both rest frames. While there is a small shift introduced by the

redshift modification, it is only makes a modest impact on the very large differences from a uniform expansion seen in the unadjusted redshifts.

It should be noted that by (9) δH_s depends on the asymptotically normalized Hubble constant, $\bar{H}_0$. Thus the vertical shift seen in Fig. 11 when the FLRW "correction" is applied is primarily due to a change in the asymptotic value, $\bar{H}_0$, as data at smaller values of cz are barely affected by the correction in (19). Table 2 gives the numerical values of the Hubble constant with adjusted and raw redshifts, along with the number of objects in each shell and the mean shell radii, making it transparent how the shift in Figure 11 arises.

As we will see in §6, the treatment of Malmquist biases is the most likely cause of the systematic difference we see in Figure 11 as compared to Fig. 3 of Wiltshire et al. (2013) for the COMPOSITE sample. To study the nature of this systematic difference we make use of the SFI++ sample, which is a subset of both the COMPOSITE and CF2 samples. SFI++ (Springob et al. 2007) consists of Tully-Fisher (1977) Relation (TFR) derived distances for 4861 field and cluster galaxies. Since Springob et al. (2007) supply the SFI++ sample with and without corrections for Malmquist biases, it makes it an ideal candidate with which to study the effects that the treatment of such biases has on the spherically averaged Hubble expansion.

To understand the effect of the Malmquist bias corrections applied by Springob et al. (2007), we calculate the monopole variation of the Hubble expansion for the SFI++ sample with and without corrections. Figure 12 shows the significant difference in δH_s between these two treatments.

Since uncorrected SFI++ data points are included in the CF2 catalogue, we can determine whether there is any systematic difference between this subsample and the remainder of the CF2 catalogue. If we take the CF2 catalogue and remove the 3625 points in common with our SFI++ sample, and repeat the analysis, we arrive at Fig. 13. We find that δH_s does not change to any statistically significant extent, by removal of this potentially biased data. This is an indication that the systematic bias present in the SFI++ raw distances – uncorrected for Malmquist bias – is likely to also be present in the rest of the CF2 data. Thus we are confident that the discrepancy seen between the monopole variation of Fig. 11 for CF2 and Fig. 3 of Wiltshire et al. (2013) for the COMPOSITE sample is a systematic issue, arising from the the treatment of Malmquist bias in the CF2 catalogue, as we discuss further in §6.

5.2 The systematic boost offset revisited

One may ask whether, despite the obvious problems with a systematic bias in the CF2 data, any of the analyses applied to the COMPOSITE can nonetheless yield meaningful results.

We applied the analysis of §4.1 in the CF2 catalogue, but found no reference frame in which χ_a^2 approaches unity, or even within the same order of magnitude. While some decrease in the variation of Hubble expansion was found for boosts in the galactic plane, the uncertainties were too large to give any statistically significant results. Such investigations must be abandoned until the bias problems in the CF2 catalogue are dealt with.

---

[10] *Cosmicflows*-2 distances retrieved 14 October 2014 from http://edd.ifa.hawaii.edu/.
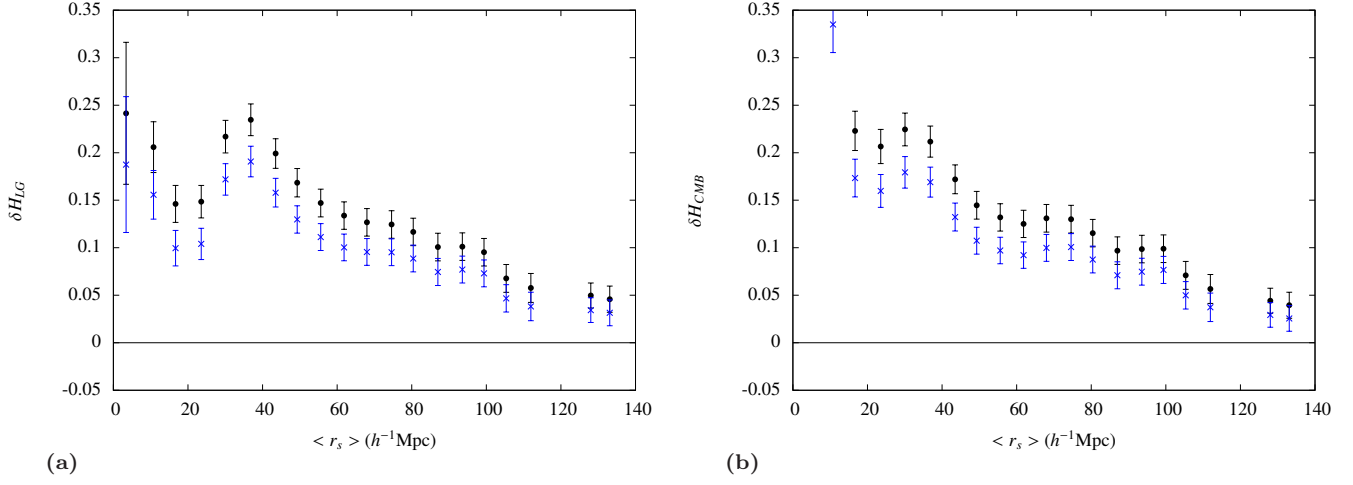
(a)

(b)

Figure 11. The monopole Hubble expansion variation for the CF2 sample without the FLRW "correction" (19) (*black* filled circles) and with the FLRW "correction" (19) (*blue* crosses) in the: **(a)** Local Group frame of reference; **(b)** CMB frame of reference.

Table 2. Hubble expansion variation in radial shells in CMB and LG frames for the CF2 data. Spherical averages (4) are computed for two different choices of shells, $r_s < r \leqslant r_{s+1}$, the second choice being labeled by primes. In each case we tabulate the number of data points per shell, the weighted mean distance, $\bar{r}_s$; the shell Hubble constants, $(H_s)_{\text{LG}}$ and $(H_s)_{\text{CMB}}$ and their associated uncertainties in the LG and CMB frames for both the raw redshifts and those adjusted with (19).

| Shell s	1	2	3	4	5	6	7	8	9	10	11
N_s	579	946	834	936	959	794	739	670	497	825	333
r_s (h^{-1}Mpc)	2.00	12.50	25.00	37.50	50.00	62.50	75.00	87.50	100.00	112.50	156.25
$\langle r \rangle_s$ (h^{-1}Mpc)	3.41	16.67	30.07	43.49	55.59	67.99	80.40	93.57	105.34	128.00	186.90
$(H_s)_{\text{CMB}}$	177.3	110.6	110.8	106.0	102.4	102.3	100.9	99.4	96.9	94.5	90.5
$(\sigma_s)_{\text{CMB}}$	10.5	1.5	1.0	0.8	0.7	0.8	0.8	0.8	0.9	0.7	0.9
$(H_s)_{\text{CMB,adjusted}}$	177.7	111.3	111.9	107.4	104.1	104.3	103.2	102.0	99.6	97.7	94.9
$(\bar{\sigma}_s)_{\text{CMB,adjusted}}$	10.5	1.5	1.1	0.8	0.8	0.8	0.8	0.8	0.9	0.7	1.0
$(H_s)_{\text{LG}}$	112.2	103.6	110.0	108.4	103.7	101.8	100.9	99.5	96.5	94.9	90.4
$(\sigma_s)_{\text{LG}}$	6.6	1.4	1.0	0.8	0.8	0.8	0.8	0.8	0.9	0.7	0.9
$(H_s)_{\text{LG,adjusted}}$	112.6	104.2	111.1	109.8	105.3	103.8	103.2	102.1	99.2	98.0	94.8
$(\sigma_s)_{\text{LG,adjusted}}$	6.7	1.4	1.1	0.8	0.8	0.8	0.8	0.8	0.9	0.7	1.0
Shell s	$1'$	$2'$	$3'$	$4'$	$5'$	$6'$	$7'$	$8'$	$9'$	$10'$	$11'$
N_s	869	867	846	989	889	777	643	648	412	625	333
r_s (h^{-1}Mpc)	6.25	18.75	31.25	43.75	56.25	68.75	81.25	93.75	106.25	118.75	156.25
$\langle r \rangle_s$ (h^{-1}Mpc)	10.76	23.54	36.85	49.29	61.86	74.59	87.01	99.37	111.95	133.10	186.90
$(H_s)_{\text{CMB}}$	126.1	109.2	109.6	103.6	101.8	102.2	99.2	99.4	95.6	94.0	90.5
$(\sigma_s)_{\text{CMB}}$	2.5	1.2	0.9	0.8	0.7	0.8	0.8	0.8	0.9	0.8	0.9
$(H_s)_{\text{CMB,adjusted}}$	126.7	110.0	110.9	105.1	103.6	104.4	101.6	102.2	98.4	97.3	94.9
$(\sigma_s)_{\text{CMB,adjusted}}$	2.5	1.2	0.9	0.8	0.8	0.8	0.8	0.8	1.0	0.8	1.0
$(H_s)_{\text{LG}}$	109.0	103.8	111.6	105.6	102.5	101.6	99.5	99.0	95.6	94.5	90.4
$(\sigma_s)_{\text{LG}}$	2.1	1.1	1.0	0.8	0.7	0.8	0.8	0.8	0.9	0.8	0.9
$(H_s)_{\text{LG,adjusted}}$	109.5	104.6	112.9	107.1	104.3	103.8	101.8	101.7	98.4	97.8	94.8
$(\sigma_s)_{\text{LG,adjusted}}$	2.1	1.1	1.0	0.8	0.8	0.8	0.8	0.8	1.0	0.8	1.0
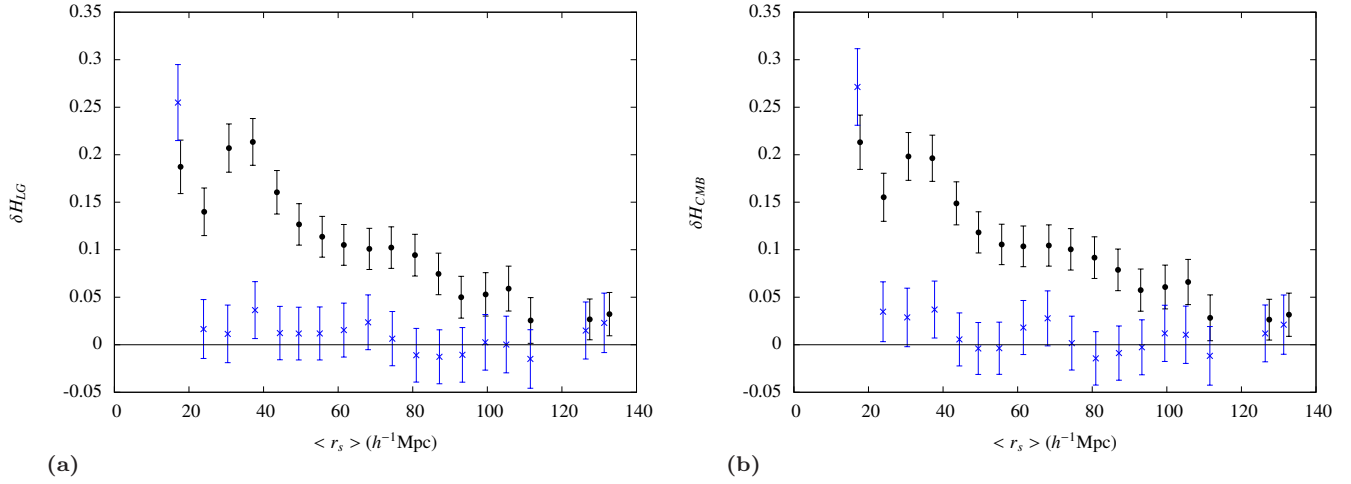
Figure 12. The monopole Hubble expansion variation for the SFI++ sample without corrections for Malmquist bias (*black* filled circles) and with corrections (*blue* crosses) in: **(a)** the Local Group frame of reference; **(b)** the CMB frame of reference.
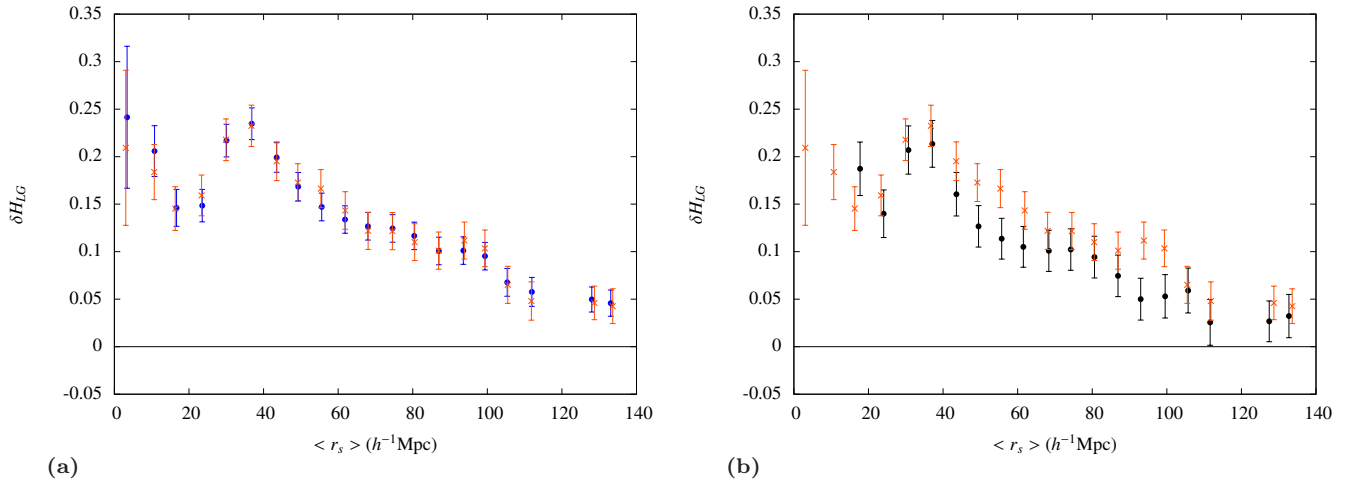

Figure 13. The monopole Hubble expansion variation in the LG frame for: **(a)** all CF2 data (*blue* filled circles) and CF2 data without SFI++ data (*red* crosses); **(b)** CF2 data without SFI++ data (*red* crosses) and SFI++ only without Malmquist corrections (*black* filled circles). For the uncorrected SFI++ data the first unprimed and primed values are not shown, these are 2.7 and 0.65 respectively.

By contrast, we found that in spite of the bias problem, the signature of a systematic boost offset studied in §3 is nonetheless evident in CF2, as this involves the *difference* of the H_s values in the LG and CMB frames from Table 2, as plotted in panels (a) and (b) of Fig. 11.

Fig. 14 shows the results of repeating the analysis used in §3, using the unadjusted CF2 distances. The best fit value for p in (15) is found to be $p = -0.83 \pm 0.17$ for unprimed and $p = -0.86 \pm 0.26$ for primed shells. Varying the shell boundaries as in §3 we find a value of $p = -0.84 \pm (0.21)_{\text{stat}} \pm (0.06)_{\text{sys}}$. However, if we compare Fig. 14 with Fig. 2, we see that there are more data points with $(H_s)_{\text{CMB}} < (H_s)_{\text{LG}}$, which do not conform to the power law (15). However, the range of distances of the shells for which this is true coincides in Fig. 2 and Figure 14, being $40\,h^{-1} \lesssim r \lesssim 60\,h^{-1}\text{Mpc}$ in the COMPOSITE sample and $30\,h^{-1} \lesssim r \lesssim 67\,h^{-1}\text{Mpc}$ in the CF2 sample. This is consistent with the hypothesis that aside from the systematic boost offset, there are structures responsible for nonlinear

deviations in the monopole Hubble expansion in the range identified in the COMPOSITE sample, but untreated biases in the CF2 catalogue have broadened the range of distances attributed to the same structures[11].

The systematic boost offset is still evident in the innermost shells of the CF2 sample. However, the fit is somewhat worse than in the COMPOSITE sample due to more data points lying in the increased range which deviates from (15). Nonetheless, we can still check if the boost offset signature is unique to the angular direction of the residual CMB dipole in the LG frame. Fig. 15 shows the value of f_p from (16) which represent the best fit parameters for a systematic boost offset for boosts of $635\,\text{km}\;\text{sec}^{-1}$ across the entire sky. The results are consistent with those found for the COMPOSITE sample, providing further evidence that this

---

[11] These data points are necessarily disregarded when we perform the logarithmic transformation to fit the power law (15), and so do not contribute to the stated uncertainties.
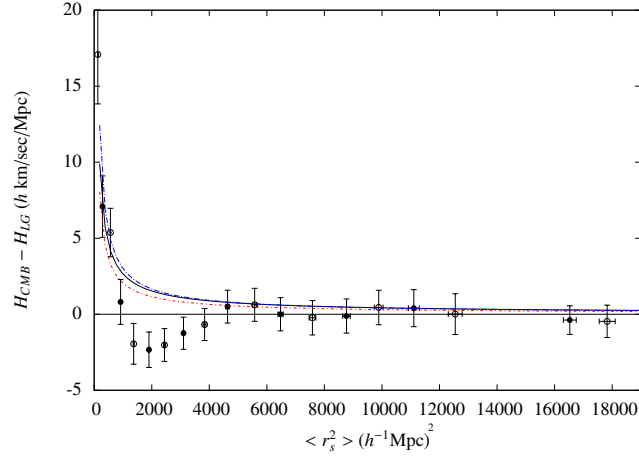
Figure 14. Best fit power law to the radial variation in the spherically averaged Hubble law in 11 shells for the CF2 galaxies data. The dashed blue curve is the best fit to primed shells only (empty circles), the dotted red curve is the best fit to unprimed shells only (filled circles) and the solid black curve is the best fit to all data points. The first data point – corresponding to unprimed shell 1 – is omitted, as it is off the scale.
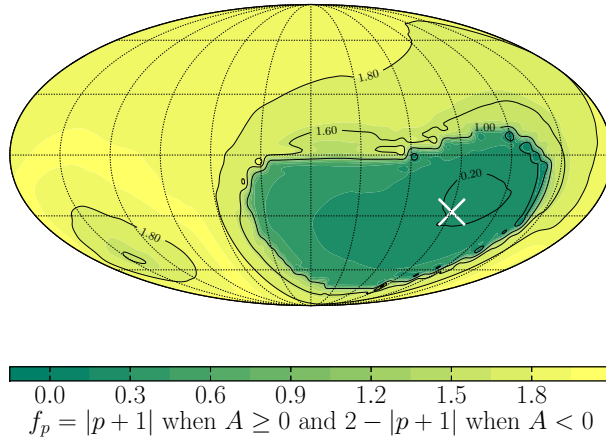

$$f_p = |p + 1| \text{ when } A \geq 0 \text{ and } 2 - |p + 1| \text{ when } A < 0$$

Figure 15. Best fit power law parameters to (15) across entire sky for boosts of $635 \, \mathrm{km \ sec^{-1}}$ from the CMB frame. The cross indicates the direction of the boost to the LG, which is also of magnitude $635 \, \mathrm{km \ sec^{-1}}$. In all figures, the galactic longitudes $\ell = 0°, 180°, 360°$ are on the right edge, centre and left edge respectively.

is indeed not a random statistical outcome but is consistent with our hypothesis.

We have also repeated the analysis of §4.3 for the best fit to a systematic boost offset from the LG frame, as shown in Figure 16. The angular directions found are consistent with our results for the COMPOSITE sample. For example, on the $200 \, \mathrm{km \ sec^{-1}}$ sky map the best fit value is in the direction $(\ell, b) = (55°, -5°)$, close to that found earlier. However, the value of $p = -0.92 \pm 0.75$ has a far greater statistical uncertainty, which may well be due to the untreated biases.

In conclusion, the greater number of data points in the CF2 catalogue may potentially yield statistically more accurate results than the COMPOSITE sample. However, at present this is prevented on account of untreated biases, to which we now turn.

6 *COSMICFLOWS*-2 MALMQUIST TREATMENT

The *Cosmicflows*-2 (CF2) and COMPOSITE catalogues deal with Malmquist distance biases in different ways. The results of the last section indicate that the treatment of such biases is crucial in establishing the actual nature of the variation of cosmic expansion below the statistical homogeneity scale. Therefore we will perform further analyses to better understand these differences.

Since the treatment of Malmquist bias is complex, we will first briefly remind the reader that in its current usage this term refers to at least three distinct biases that affect the average derived distance of a galaxy cluster:

(i) *Selection bias* This is the simple homogeneous systematic error that results from using objects in a magnitude limited sample (Malmquist 1922). If the galaxies in a cluster have a
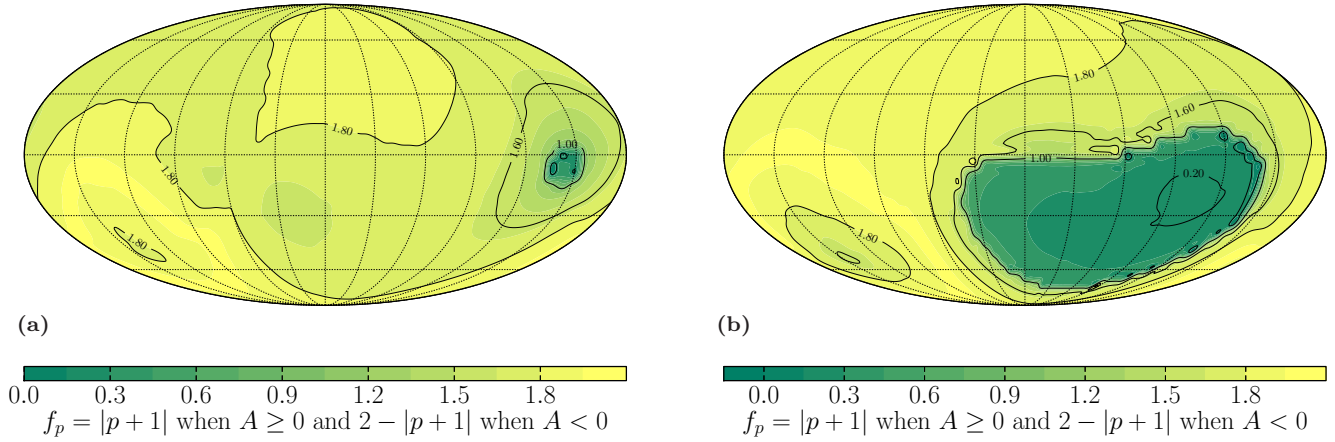
Figure 16. Best fit parameters to (15) using unadjusted CF2 data for boosts from the LG of: **(a)** $450 \, \text{km} \, \text{sec}^{-1}$; **(b)** $200 \, \text{km} \, \text{sec}^{-1}$. In all figures, the galactic longitudes $\ell = 0°, 180°, 360°$ are on the right edge, centre and left edge respectively.

true average flux $\langle F \rangle_0$, since we can only observe the brightest of these, we measure a biased average flux $\langle F \rangle_b > \langle F \rangle_0$, with the effect of bias increasing with distance. Thus the luminosity distance $d_L{}^2 = \langle L \rangle_0 / (4\pi \langle F \rangle)$ will be less when using a biased average flux, where $\langle L \rangle_0$ is the average luminosity as determined from standard candles calibrated using nearby objects.

(ii) *Homogeneous distribution bias* This is a systematic average error for objects around the same derived distance which can be understood in terms of statistical scatter. If we assume a standard Gaussian scatter, σ, in the derived distances about an estimated mean, then – since the radial number density grows as $N(r) \propto r^3$ – there are more objects with true distances larger than the estimated distance, than smaller. Thus at a given derived distance distance, more galaxies will have been scattered by the errors down from larger true distances than up from smaller ones (Lynden-Bell et al. 1988; Hanski 1999). This is equivalent to giving more statistical weight to more distant values and thus the probability distribution for the true distance is no longer Gaussian along the line of sight, centred on the measured distance, instead being skewed towards greater distances. For the special case of a constant density distribution, for example, one arrives at Eddington's formula (Eddington 1914)

$$E(\mu_{\text{true}}|\mu_{\text{der}}) = \mu_{\text{derived}} + 1.382\sigma^2, \qquad (20)$$

where $\mu \equiv 5 \log r + 25$ is the standard distance modulus when r is given in Mpc.

(iii) *Inhomogeneous distribution bias* The inhomogeneous bias is analogous to the homogeneous one in that it involves a systematic error in the statistical scatter of objects due to their distribution in 3–dimensional space. However, in this case it arises as number counts are higher in regions of greater density, resulting in a systematic scatter of measurements out of higher into lower density regions (Strauss & Willick 1995). Thus the inhomogeneous effect is very sensitive to large variations in large-scale structure along the line of sight. Failure

to account for this type of bias can give spurious infall signatures onto high density regions. This bias is of course far more difficult to account for, requiring accurate density fields for structure along the line of sight for each observation.

The selection and homogeneous distribution biases typically lead to underestimates of distances that increase as the distance grows, so that a plot of $\mu_{\text{true}} - \mu_{\text{der}}$ versus redshift has a positive slope. However, the inhomogeneous distribution bias can lead to the opposite effect. For example, Feast (1987) shows that in applying the TFR method when the spatial density of objects at a given 21cm line width is constant, then the required Malmquist correction is the classical one given by Eddington (1914). However, when this is not the case it is possible to obtain overestimated distances (Feast 1987).

Moreover, regardless of the details of any inhomogeneous matter distribution, one just needs the spherically averaged $N(r)$ to decrease sufficiently quickly for the direction of scatter in the standard homogeneous distribution bias to be reversed, giving overestimated distances.

Some biases can be dealt with in the data reduction. In particular, by applying the inverse TFR method rather than the direct method one can in principle effectively eliminate the selection bias (Schechter 1980), leaving only a considerably smaller bias (Willick 1994, 1995).

In the CF2 catalogue Tully et al. (2013) use an inverse TFR procedure to reduce the selection bias only, stating that only a small subsequent correction for residual bias is required. In particular, they "make no adjustments for the distribution Malmquist effects" in their reported CF2 distances (Tully et al. 2013). Their calibration carried out for this relation follows the procedures of Tully & Pierce (2000), Courtois & Tully (2012) and Sorce et al. (2013).

On the other hand, the SFI++ catalogue (Springob et al. 2007) which forms the major part of the COMPOSITE sample includes corrections to account for all homogeneous and inhomogeneous biases in their

data. In their view their own treatment of the homogeneous and inhomogeneous distribution biases was "straightforward", given access to a reconstruction of the local density field (Erdoğdu et al. 2006). However, they stated that their treatment of the selection bias was *ad hoc* because the selection criteria used are designed to mimic the observational properties of the survey as closely as possible, and so are very inhomogeneous. They therefore provided both the raw and corrected distances, should other researchers adopt alternative methods for dealing with selection bias.

The COMPOSITE sample incorporates the SFI++ distances with the Malmquist bias corrections of Springob et al. (2007), whereas CF2 uses a subset of uncorrected SFI++ distances. While we do not have the data to independently repeat any of the bias corrections, we are able to test the consistency of assumptions made by Tully et al. (2013).

6.1 The SFI++ subsamples of CF2

Tully et al. (2013) find that for 2071 common points between their own survey and the SFI++ survey (excluding 5 points judged to be "bad") there was a "correction" of the form

$$\Delta\mu_1 = 0.492(\pm 0.011) + 0.000031(\pm 0.000002)cz_{\rm LS} \qquad (21)$$

where $\Delta\mu_1 \equiv \mu_{cf2} - \mu_{sfi}$, μ_{cf2} is the CF2 distance modulus with the zero point established by Courtois & Tully (2012), and μ_{sfi} is the Springob et al. (2007) unadjusted modulus with a nominal zero point consistent with $H_0 = 100h$ km sec^{-1} Mpc^{-1}, and $z_{\rm LS}$ is the raw[12] redshift in the rest frame of the Local Sheet, which is close to the Local Group frame (Tully et al. 2008). We independently confirm the slope in (21) using the appropriate zero point[13], and plot this in Figure 17. Note that this comparison is made for a subset of the SFI++ sample, henceforth SFI++A, consisting only of objects that are common between the SFI++ survey and the independently obtained CF2 distances. In their final analysis Tully et al. (2013) use averages of the CF2 and SFI++ distances with double weight given to the CF2 distances.

The intercept in Fig. 17(a) simply reflects the different normalizations of H_0 in the two samples. The positive slope of the linear relation in Fig. 17(a) is consistent with the CF2 inverse TFR distances having a correction that accounts for the raw SFI++ distances being increasingly underestimated due to the Malmquist selection bias.

Let us now compare (17) to the Malmquist correction used by Springob et al. (2007) in the SFI++ sample, and subsequently adopted in the COMPOSITE sample. We repeat the analysis of Fig. 17 on the SFI++A subsample but now using the distances as corrected by Springob et al. (2007), which include both selection and distribution bias corrections. We find a linear relationship of the form

$$\Delta\mu_2 = 0.0356(\pm 0.0063) - 0.000012(\pm 0.000001)cz_{\rm LS} \qquad (22)$$

where $\Delta\mu_2 \equiv \mu_{sfi,corrected} - \mu_{sfi}$ is the difference in distance moduli between the corrected and raw distances. The data and best fit line are displayed in Fig. 17.

We can immediately see from (21) and (22) that there is a significant difference between the corrections. For small redshifts the Springob et al. (2007) correction is positive indicating raw distances are underestimated, while for large redshifts the correction is negative indicating raw distances are overestimated. Adjusting the intercept of Fig. 17(a) to zero we then find a hierarchy $\mu_{sfi,corrected} < \mu_{sfi} < \mu_{cf2}$ in the limit of large redshifts. This is consistent with the observation of Watkins & Feldman (2015) that: "the distances are systematically larger in the *Cosmicflows*-2 catalogue [than in the COMPOSITE catalogue] due to a different approach to bias correction".

The fact that Fig. 22(b) has a negative slope means that the dominant correction cannot arise solely from selection and homogeneous distribution biases, since as noted above both of these effects underestimate true distances. The difference therefore must be due to the treatment of the inhomogeneous distribution bias, which Springob et al. (2007) have included but Tully et al. (2013) have not.

The 1970 points in the SFI++ sample that are not also contained in the original CF2 survey, henceforth SFI++B, have been incorporated into CF2 without using their correction (21). This data covers a larger range, up to redshifts of almost $z = 0.1$, whereas SFI++A only covers up to $z = 0.06$. Tully et al. (2013) state that these points, if corrected using (21) cause a "highly significant decrease in the Hubble parameter with increasing velocity". We independently verified this result[14]. Thus Tully et al. (2013) do not adjust these distances, instead claiming that they are of a different nature altogether, the main difference being that these consist of cluster samples from a different survey (Dale et al. 1999a,b). In these samples rotation information for the galaxies was obtained from optical spectroscopy rather than the standard 21cm Hydrogen line widths. However, Tully et al. state that "it is not clear to us why this component of SFI++ does not manifest the selection Malmquist bias".

Since it appears two halves of the SFI++ have been incorporated into CF2 in different ways we determine whether they do actually show different characteristics. Considering $\Delta\mu_2$ for SFI++B we find a correction

$$\Delta\mu_2 = 0.0417(\pm 0.0061) - 0.000012(\pm 0.000001)cz_{\rm LS} \qquad (23)$$

which has an intercept consistent within uncertainties and identical slope to (22). Since there is no apparent difference

---

[12] The FLRW adjustment (19) is not applied.
[13] The intercept in (21) is determined by a scaling of the data so we are not interested in independently confirming this for our investigation.
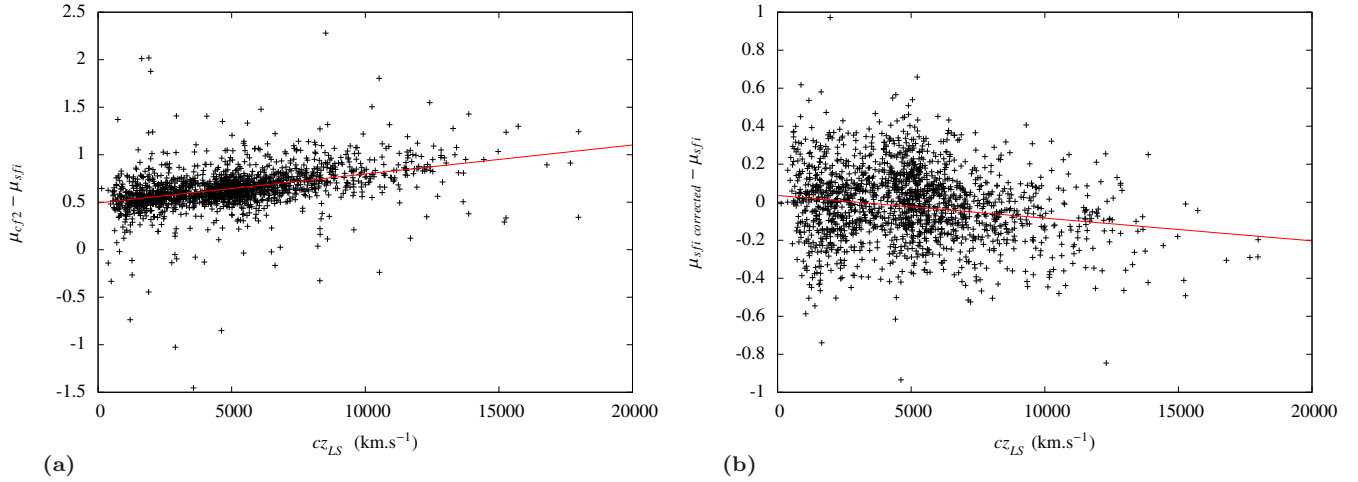
[14] A decreasing Hubble constant below the scale of statistical homogeneity is, to a limited extent, what is expected from the analysis of Wiltshire et al. (2013). Thus trends which appear anomalous as compared to a standard FLRW expectation should not automatically be regarded as a signal of unaccounted observational bias. However, there are also systematic differences that occur when binning in redshift, as in Figs. 18 and 19, as opposed to binning in distance with (4), so careful analysis is required to make sense of the different approaches. The direct calculation of the Hubble parameters in each bin is also different to that described in (4), as it is not clear which method Tully et al. (2013) use. We apply a simple weighted average of cz_i/r_i values and obtain consistent results (although as no values are tabulated by Tully et al. (2013) we can only verify by inspection).

Figure 17. The difference in distance modulus between: **(a)** the CF2 and raw SFI++ distances for the points common to the two surveys, as presented in Tully et al. (2013). **(b)** the SFI++ raw and corrected catalogues for the same points.

using this test we repeat a similar analysis to that performed by Tully et al. (2013, Fig. 10) to test for bias. This analysis is based on the fact that selection bias is manifest by an increase in the Hubble parameter with redshift (Teerikorpi 1993), for data binned by redshift. In Figs. 18 and 19 we repeat the analysis of the Hubble constant in redshift bins performed by Tully et al. (2013) for the subsets of interest, and compare the results.

In Fig. 18 we produce plots equivalent to Tully et al. (2013, Fig. 10) for the SFI++A and SFI++B subsamples, both using raw distances. We subsequently find that the difference in the Hubble constant in individual redshift bins for the SFI++A and SFI++B ranges from 0.03σ to 1.8σ in individual bins. The weighted mean of these differences is 0.84σ, and thus we do not see a significant difference between SFI++A and SFI++B.

We note that two mutually consistent halves of the SFI++ sample have been incorporated into CF2 in different ways, but are uncertain as to what impact this may have on the full CF2 catalogue.

On the other hand the COMPOSITE sample (Feldman et al. 2010) uses a much larger subset of SFI++ distances corrected for Malmquist biases (after rejection of outliers). While it appears that there may be inconsistencies in the inclusion of the raw SFI++ distances into the CF2 catalogue, it is possible that the SFI++ corrected distances are subject to systematic error also. It is for this reason that Springob et al. (2007) included both corrected and raw distances, to allow others to take on the challenging task of Malmquist bias corrections. However, the CF2 catalogue is only corrected for selection bias, leaving the correction of the distribution homogeneous and inhomogeneous Malmquist biases as a task for the user.

As another test of differences between the SFI++A and SFI++B subsamples we have also repeated the analysis of Fig. 18, but now to compare the raw and corrected distances within each subsample. Fig. 19 shows the comparison for the SFI++B subsample. In this case the difference in the Hubble parameters in each bin vary from a minimum 0.01σ to a maximum 2.2σ in individual bins, with a weighted mean difference of 1σ. For SFI++A the difference in the Hubble

parameters vary from 0.05σ to 1.9σ in individual bins, with a weighted mean difference of 1.0σ. Thus again we do not see a significant difference between the subsamples.

It may appear surprising that the raw and uncorrected data only differ by 1σ on average when binned by redshift. However, once inhomogeneous Malmquist bias is accounted for the sign of the correction is different at large redshifts as compared to low redshifts, meaning that for an intermediate range the correction is small. The approach by Tully et al. (2013) of binning in redshift is not an appropriate one to use when performing a parameter minimization that involves boosts to rest frames in which the redshift is changed, as in §3–4. Rather we followed Wiltshire et al. (2013) in binning by distance. This led to differences between the raw and corrected data sets which are statistically much more marked than is evident if one bins by redshift.

7   DISCUSSION

We have investigated the extent to which it is possible to define a standard of cosmological rest based on a frame in which variation of the spherically averaged (monopole) nonlinear Hubble expansion is a minimum. Such averages do not make any assumptions about the geometry of space below the scale of statistical homogeneity ($\lesssim 100\,h^{-1}\mathrm{Mpc}$) of the sort implicit in the standard peculiar velocities framework, which assumes a Euclidean spatial geometry.

We studied the systematic variation that arises when an arbitrary boost is made from a frame of reference in which the spherically averaged Hubble expansion is most uniform. We found that such a systematic variation is indeed detected to a statistically significant extent between the CMB and LG frames of reference, using the COMPOSITE sample (Watkins et al. 2009; Feldman et al. 2010).

This supports the proposition made by Wiltshire et al. (2013) that the Local Group frame may be a more suitable cosmic rest frame than the standard CMB frame, and that consequently a significant fraction of the CMB dipole may be nonkinematic. This proposition can be directly tested in numerical ray–tracing simulations (Bolejko et al. 2015). Given
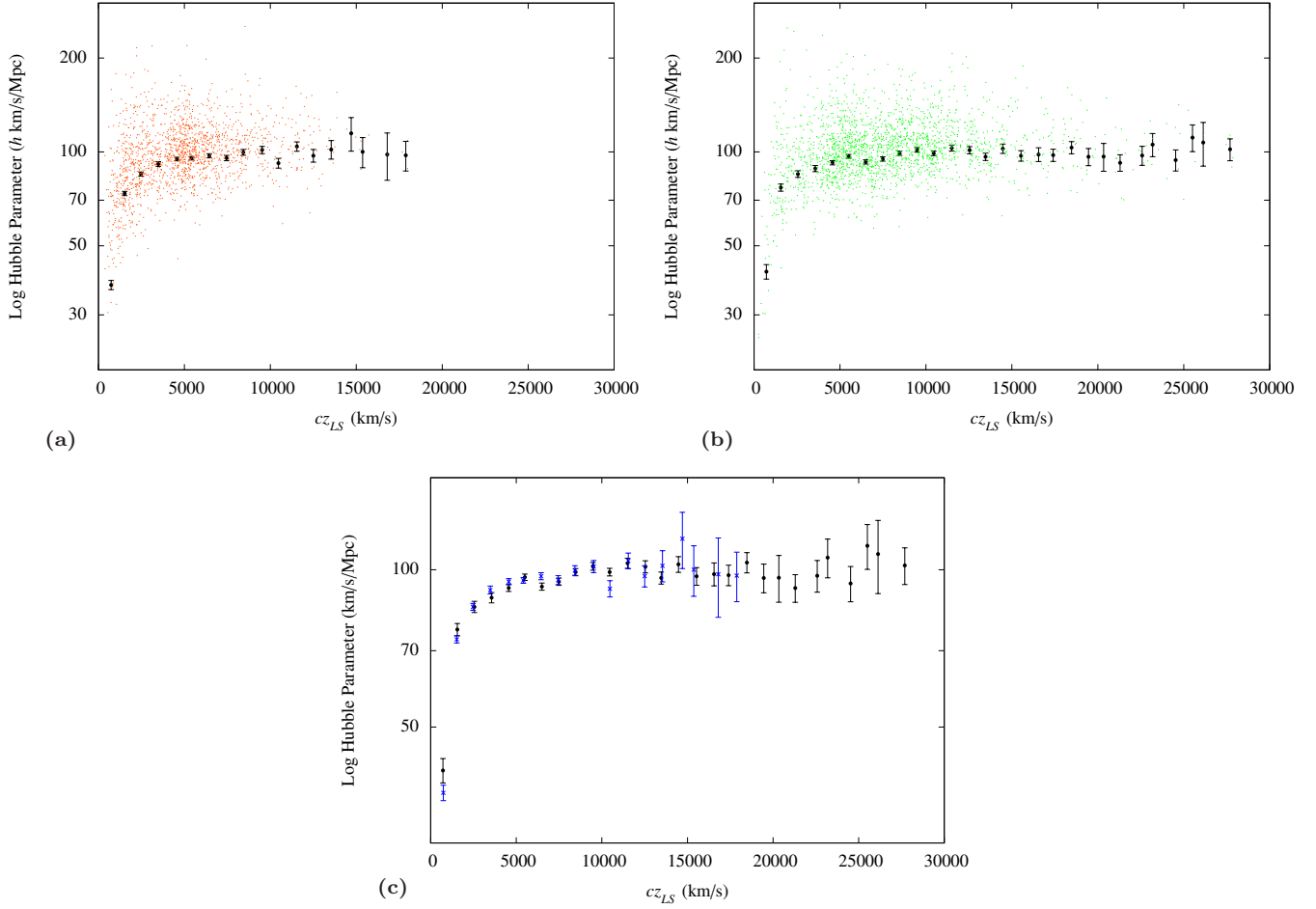
Figure 18. The Hubble parameter, $H_i = cz_i/r_i$, computed for each individual data point (coloured points) and from averaging in $1000 \, \mathrm{km \, sec^{-1}}$ bins (black points) using the: **(a)** SFI++A subsample; **(b)** SFI++B subsample. **(c)** Comparison of the averaged points in (a) and (b) with blue crosses being from SFI++A and black filled circles from SFI++B.
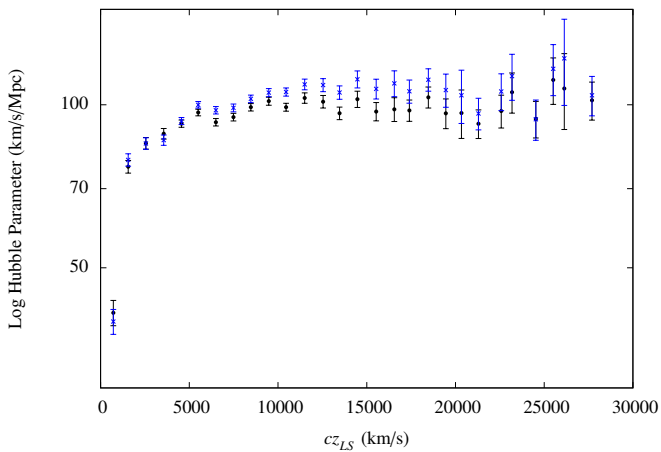

Figure 19. The Hubble parameter, $H_i = cz_i/r_i$, computed from averaging in $1000 \, \mathrm{km \, sec^{-1}}$ bins using the SFI++B subsample with corrections for Malmquist bias (blue crosses) and without corrections (black filled circles).

that the CMB frame is still the de facto choice for the cosmic rest frame, this conclusion would have a far reaching impact for many areas of cosmology (Wiltshire et al. 2013), including the question of large angle CMB anomalies (Ade et al. 2014b).

We extended the analysis to search for an improvement on the Local Group as the standard of rest, using a variety of tests. Very large boosts from the LG frame can be excluded if we simultaneously demand that while the residual variation of the Hubble expansion should be small in the regime of nonlinear expansion at small distances, a clear signature of an emerging linear Hubble law should also be found at larger distances. However, we found that in applying all possible tests there is still freedom to perform quite large boosts close to the plane of the galaxy, presumably because the lack of data in the Zone of Avoidance leads to a lack of suitable constraints there. This hypothesis could potentially be tested on simulated data using exact solutions of Einstein's equations (Bolejko et al. 2015).

Since our conclusions depend on the COMPOSITE sample, we repeated our analysis insofar as it was possible for the recently released *Cosmicflows*-2 sample (Tully et al. 2013). This catalogue of 8162 galaxy, groups and cluster distances is considerably larger than the the COMPOSITE

sample, and potentially could provide more accurate results although it is also of course limited in the Zone of Avoidance.

We have found very significant differences in the results for the CF2 and COMPOSITE samples, as can be seen by comparing Fig. 3 of Wiltshire et al. (2013) and our Fig. 11. These result from differences in the treatment of the Malmquist bias between the SFI++ catalogue and CF2 catalogue, as previously noted by Watkins & Feldman (2015). We also found apparent inconsistencies in the manner of inclusion of subsamples of the SFI++ catalogue into the CF2 catalogue, with respect to the treatment of Malmquist bias. More significantly, since the reported CF2 distances do not include corrections for the inhomogeneous distribution Malmquist bias they may be of limited use until such corrections are applied.

The conclusions of Wiltshire et al. (2013) are dependent on the treatment of the Malmquist bias in the SFI++ catalogue being accurate, as this constitutes the largest part of the COMPOSITE sample. Naturally, one might question whether any systematic procedure of Springob et al. (2007) could somehow spuriously lead to an unusually uniform Hubble expansion in the LG frame through an error in the Malmquist bias procedure.

We find no grounds for this. In particular, our analysis shows that the difference between the CMB and LG frames has the distinctive signature of a systematic boost offset (12) noted by Wiltshire et al. (2013). Nothing in the Malmquist bias correction procedure of Springob et al. (2007) could obviously introduce this signature through a systematic error. Their analysis does not single out the LG, or LS, frame in any way; indeed all their redshifts are referred to the CMB frame. Furthermore, although the remaining bias means that the CF2 sample is currently unusable from the point of view of determining a frame of minimum spherically averaged Hubble expansion variation in the nonlinear regime, we observed in §5.2 that the *difference* of the CMB and LG frame spherical averages nonetheless still shows the signature of the systematic boost offset in the CF2 data.

Since the boost offset is detectable in the independently reduced CF2 data, it cannot be an artefact of the Malmquist bias treatment of Springob et al. (2007). Furthermore, the departure of the nonlinear expansion from the boost power law (15) that is seen when comparing Figs. 2 and 14 is precisely what is to be expected if there are additional unaccounted uncertainties in individual CF2 distances as compared to the SFI++ ones: the distance range of structures associated with the nonlinear expansion is broadened.

If Fig. 11 was based on accurate distances, it would imply that the Hubble expansion in all frames of reference is far less uniform than might reasonably be expected in any viable cosmological model; in particular, there is a monopole or "Hubble bubble" variation of order 15–20% in the range $20 < r < 60\,h^{-1}\mathrm{Mpc}$ in the CF2 sample, as compared to 4–5% in the COMPOSITE sample. The largest "Hubble bubble" variation that has ever been claimed on such scales using more accurate Type Ia supernovae distances in the CMB frame is $6.5 \pm 2.2\%$ (Zehavi et al. 1998).

Tully et al. (2013) chose not to correct for the distribution biases, as they wished to separate "the issues of distance measurements and velocity field inferences". Indeed, in the peculiar velocities approach the distribution bias may be much less significant. In new work

Hoffman, Courtois & Tully (2015) use the CF2 catalogue to reconstruct large scale structure by means of the Wiener filter and constrained realizations of Gaussian fields assuming a WMAP constrained ΛCDM model as a Bayesian prior. They observe that "the Malmquist bias introduces a spurious strong monopole term into the reconstructed velocity field but is expected to hardly affect the bulk velocity which is associated with the dipole of the velocity field" (Hoffman et al. 2015). This would appear to be the counterpart of the large monopole we observe in Fig. 11 in our analysis. Hoffman et al. (2015) corrected for the bias but noted that the bulk velocity analysis is "virtually unaffected by the Malmquist bias".

In our case, residual bulk flows on scales $\gtrsim 100\,h^{-1}\mathrm{Mpc}$ may be an artefact of using the CMB rest frame as the standard when it does not coincide with the frame of minimum Hubble expansion variation. In particular, Wiltshire et al. (2013) observed that a dipole fit to the Hubble expansion gave a dipole magnitude which increased to a residual value in the outer shells as the shell radius was increased in the CMB frame, but not in the LG frame. Thus large scale bulk flows are not our primary interest. Rather, we are interested in detecting the systematic monopole variation (12). In distinction to the peculiar velocity approach our method by necessity is sensitive to a monopole bias. In fact, our method of binning in radial shells by distance with an anchoring to $\bar{H}_0$, is particularly sensitive to any distribution bias which follows from a number density, $N(r)$, with strong gradients. The bias effect in Fig. 11 can be largely reproduced by applying a uniform Hubble law to the CF2 redshifts, adding Gaussian scatter to create a mock distance catalogue, and then applying our binning strategy (R. Watkins, private communication).

It may be possible to construct the $635\,\mathrm{km}\,\mathrm{sec}^{-1}$ velocity attributed to the LG within the ΛCDM model, as has recently been claimed by Hess & Kitaura (2016) who used constrained N-body simulations and nonlinear phase space reconstructions to arrive at a value $v_{\mathrm{LG}} = 685 \pm 137\,\mathrm{km}\,\mathrm{sec}^{-1}$. However, this itself does not constitute a proof of the standard kinematic interpretation, but rather a verification within the 20% uncertainty of a computer simulation. The ΛCDM model is certainly phenomenologically very successful, and any competing model can only be viable insofar as many of its predictions are close to the standard model, as is the case, for example, in the timescape cosmology (Wiltshire 2009; Smale & Wiltshire 2011; Smale 2011; Nazer & Wiltshire 2015). What is important in testing the standard model is to seek observations which are not expected in its framework. Although the signature of the systematic monopole boost offset (12) between the CMB and LG frames should be checked in ΛCDM simulations, it is not an observation that should obviously arise if all cosmological redshifts arise purely from a FLRW geometry plus local Lorentz boosts at any scale. Properly characterizing and determining a frame of minimum nonlinear Hubble expansion variation is therefore a fundamental question open to more precise observational tests in future.

In conclusion, if we make no assumptions about our own standard of rest, but make arbitrary boosts at our location, then analysis of the COMPOSITE sample yields a degenerate set of frames for which the variation of the spherically averaged nonlinear Hubble expansion is minimized. This set

of candidate cosmic rest frames includes the LG frame but excludes the CMB frame, as was already established with very strong Bayesian evidence by Wiltshire et al. (2013). The degeneracy in defining the minimum variation frame is associated with a freedom to perform boosts in the plane of the galaxy without changing the statistical significance of the fit, probably due to the lack of constraining data in the Zone of Avoidance.

The larger CF2 sample may potentially tighten the constraints on the definition of the frame of minimum nonlinear expansion variation. However, it is first necessary to reduce the data in the manner of Springob et al. (2007) to remove the inhomogeneous distribution bias which appears to be the source of the large discrepancies found in §5. The Universe is very inhomogeneous below the scale of statistical homogeneity, and ironically it is only once the biases associated with such inhomogeneities are removed that a picture of a remarkably uniform average Hubble expansion actually emerges.

A careful treatment of inhomogeneous Malmquist bias is therefore key for the future progress of our understanding of the nature of cosmic expansion as the surveys grow ever larger.

REFERENCES

Abazajian K.N. *et al.*, 2009, ApJS, 182, 543.

Ade P.A.R., *et al.*, 2014a, A&A, 561, A97.

Ade P.A.R., *et al.*, 2014b, A&A, 571, A23.

Agarwal S., Feldman H. A., Watkins R., 2012, MNRAS, 424, 2667.

Aghanim N. *et al.*, 2014, A&A, 571, A27.

Atrio-Barandela F., 2013, A&A, 557, A116.

Bilicki M., Peacock J.A., Jarrett T.H., Cluver M.E., Steward L., 2014, arXiv:1408.0799

Bolejko K., Nazer M.A., Wiltshire D.L., 2015, arXiv:1512.07364

Bondi H., 1947, MNRAS, 107, 410

Buchert T., 2000, Gen. Relativ. Grav., 32, 105.

Buchert T., 2008, Gen. Relativ. Grav., 40, 467.

Carrick J., Turnbull S.J., Lavaux G., Hudson M.J., 2015, MNRAS, 450, 317.

Clowes R.G., Harris K.A., Raghunathan S., Campusano L.E., Soechting I.K., Graham M.J., 2013, MNRAS, 429, 2910.

Courtois H.M., Tully R.B., 2012, ApJ, 749, 174.

Dale D.A., Giovanelli R., Haynes M.P., Hardy E., Campusano L.E., 1999a, AJ, 118, 1468.

Dale D.A., Giovanelli R., Haynes M.P., Campusano L.E., Hardy E., 1999b, AJ, 118, 1489.

Eddington A.S., 1914, *Stellar Movements and the Structure of the Universe*, (Macmillan, London)

Einasto J., 2014, arXiv:1410.6932.

Ellis G.F.R., 1984, in Bertotti B., de Felice F., Pascolini A., eds, *General Relativity and Gravitation*, (Reidel, Dordrecht) pp. 215–288.

Erdoğdu P., Lahav O., Huchra J.P., Colless M., Cutri R.M., Falco E., George T., Jarrett T., 2006, MNRAS, 373, 45.

Feast M.W., 1987, The Observatory, 107, 185

Feldman H.A., Watkins R., Hudson M.J., 2010, MNRAS, 407, 2328.

Fixsen D.J., Cheng E.S., Gales J.M., Mather J.C., Shafer R.A., Wright E.L., 1996, ApJ, 473, 576.

Forero–Romero J.E., Hoffman Y., Gottlöber S., Klypin A., Yepes G., 2009, MNRAS, 396, 1815.

Freedman W.L., Madore B.F., Gibson B.K., *et al.*, 2001, ApJ, 553, 47

Gorski K.M., Hivon E., Banday A.J., Wandelt B.D., Hansen F.K., Reinecke M., Bartelman M., 2005, ApJ, 622, 759.

Gott J.R., Juric M., Schlegel D., Hoyle F., Vogeley M., Tegmark M., Bahcall N., Brinkmann J., 2005, ApJ, 624, 463.

Hanski M., 1999, in Impey C., ed, *International Symposium on Astrophysics Research and Science Education*, (University of Notre Dame Press) pp. 242-246.

Hess S., Kitaura F.S., 2016, MNRAS, in press; arXiv:1412.7310.

Hoffman Y., Courtois H.M., Tully R.B., 2015, MNRAS, 449, 4494.

Hogg D.W., Eisenstein D.J., Blanton M.R., Bahcall N.A., Brinkmann J., Gunn J.E., Schneider D.P., 2005, ApJ, 624, 54.

Hoyle F., Vogeley M.S., 2002, ApJ, 566, 641.

Hoyle F., Vogeley M.S., 2004, ApJ, 607, 751.

Kashlinsky A., Atrio-Barandela F., Ebeling H., Edge A., Kocevski D., 2010, ApJ, 712, L81.

Kass R.E., Raftery A.E., 1995, J. Am. Statist. Assoc. 90, 773.

Lavaux G., Afshordi N., Hudson M.J., 2013, MNRAS, 430, 1617.

Lemaître G., 1933, Ann. Soc. Sci. Bruxelles A, 53, 51 [English translation in Gen. Relativ. Grav., 29, 641 (1997)].

Li N., Schwarz D.J., 2008, Phys. Rev. D, 78, 083531.

Lynden-Bell D., Faber S.M., Burstein D. *et al.*, 1988, ApJ, 326, 19.

Ma Y.Z., Scott D., 2013, MNRAS, 428, 2017.

Malmquist K.G., 1922, Medd. Lund Astron. Obs. Ser. II, No. 22.

McClure M.L., Dyer C.C., 2007, New Astron. 12, 533.

McKay J.H., 2015, *"The cosmological rest frame,"* MSc thesis, University of Canterbury.

Nazer M.A., Wiltshire, D.L., 2015, Phys. Rev. D, 91, 063519.

Nusser A., Davis M., 2011, ApJ, 736, 93.

Pan D.C., Vogeley M.S., Hoyle F., Choi Y.Y., Park C., 2012, MNRAS, 421, 926.

Rubart M., Schwarz D.J., 2013, A&A, 555, A117

Schechter P., 1980, AJ, 85, 801.

Scrimgeour M. *et al.*, 2012, MNRAS, 425, 116.

Smale P.R., 2011, MNRAS, 418, 2779

Smale P.R., Wiltshire D.L., 2011, MNRAS, 413, 367

Sorce J.G. *et al.*, 2013, ApJ, 765, 94.

Springob C.M., Masters K.L., Haynes M.P., Giovanelli R., Marinoni C., 2007, ApJS, 172, 599; [Err., 2009, ApJS, 182, 474].

Strauss M.A., Willick J.A., 1995, Phys. Rep. 261, 271.

Szekeres P., 1975, Commun. Math. Phys. 41, 55.

Teerikorpi P., 1993, A&A, 280, 443

Titterington D., Halliday A., 1979, Chem. Geol., 26, 183

Tolman R.C., 1934, Proc. Nat. Acad. Sci. 20, 169.

Tully B.R., Fisher J.R., 1977, A&A, 54, 661.

Tully R.B., Pierce M.J., 2000, ApJ, 533, 744.

Tully B.R., Shaya J.E., Karachentsev D.I., Courtois M.H., Kocevski D.D., Rizzi L., Peel A., 2008, ApJ, 676, 184.

Tully R.B. *et al.*, 2013, AJ, 146, 86.

Turnbull S.J., Hudson M.J., Feldman H.A., Hicken M., Kirshner R.P., Watkins R., 2012, MNRAS, 420, 447.

Watkins R., Feldman H.A., 2015, MNRAS, 447, 132.

Watkins R., Feldman H.A., Hudson M.J., 2009, MNRAS, 392, 743.

Willick J.A., 1994, ApJS, 92, 1.

Willick J.A., Courteau S., Faber S.M., Burstein D., Dekel A., 1995, ApJ, 446, 12.

Wiltshire D.L., 2007a, New J. Phys. 9, 377.

Wiltshire D.L., 2007b, Phys. Rev. Lett., 99, 251101.

Wiltshire D.L., 2008, Phys. Rev. D, 78, 084032.
Wiltshire D.L., 2009, Phys. Rev. D, 80, 123512.
Wiltshire D.L., 2011, Class. Quantum Grav., 28, 164006.
Wiltshire D.L., Smale P.R., Mattsson T., Watkins R., 2013, Phys. Rev. D, 88, 083529.
York D., 1969, Earth Planet. Sci. Lett., 5, 320
York D., Evensen N.M., Martnez M.L., Delgado J.D. 2004, Am. J. Phys., 72, 367
Zehavi I., Riess A.G., Kirshner R.P., Dekel A., ApJ, 503, 483 (1998).

APPENDIX A: LINEAR REGRESSION VIA TOTAL LEAST SQUARES

Consider a general linear model with errors in both the dependent and independent variables. We can express such a model as

$$y_i = \beta_0 + \beta_1 x_i$$
$$(Y_i, X_i) = (y_i, x_i) + (e_i, u_i) \tag{A1}$$

where (Y_i, X_i) are the observations, (y_i, x_i) are the true values and (e_i, u_i) are the measurement errors. We assume the measurement errors to be normally distributed with a covariance matrix

$$\Sigma = \begin{pmatrix} \sigma_{ee} & \sigma_{eu} \\ \sigma_{ue} & \sigma_{uu} \end{pmatrix}. \tag{A2}$$

We can extend this analysis to allow for different errors at each point, which we will refer to as weights given by $\omega(X_i) = 1/\sigma_{uu,i}$, $\omega(Y_i) = 1/\sigma_{ee,i}$ and the correlation coefficient between the errors given by $\gamma_i = \sigma_{eu,i}\sqrt{\omega(X_i)\omega(Y_i)}$.

To carry out a least squares minimization we must calculate the statistical distance from an observation to the true value. In standard least squares this distance is the vertical distance from the data point to the model, since the values of the independent variable are assumed to be exact. Now, in the most simple case we would have the squared Euclidean distance from the observed data points to true value in the model as

$$[Y_i - (\beta_0 + \beta_1 x_i)]^2 + (X_i - x_i)^2 = e_i^2 + u_i^2 \tag{A3}$$

but if the variances of e_i and u_i are different from unity this statistical distance becomes $\sigma_{ee}^{-1}e_i^2 + \sigma_{uu}^{-1}u_i^2$, and if these variances are correlated we must use the covariance matrix to give the "statistical" distance

$$(Y_i - \beta_0 - \beta_1 x_i, X_i - x_i)\Sigma^{-1}(Y_i - \beta_0 - \beta_1 x_i, X_i - x_i)^\intercal. \tag{A4}$$

In our case we must determine the values of β_0 and β_1 that minimize (A4). That is, we must find the values $(\hat{x}_i, \hat{y}_i)$ and $(\hat{\beta}_0, \hat{\beta}_1)$ that minimize this sum for the given observations. First we fix the x_i values by treating them as unknown constants in a standard linear regression of the form

$$\begin{bmatrix} Y_i - \beta_0 \\ X_i \end{bmatrix} = \begin{bmatrix} \beta_1 \\ 1 \end{bmatrix} x_i + \begin{bmatrix} e_i \\ u_i \end{bmatrix} \tag{A5}$$

for which the generalized least squares estimator gives

$$\hat{x}_i = [(\beta_1, 1)\Sigma^{-1}(\beta_1, 1)^\intercal](\beta_1, 1)\Sigma^{-1}(Y_i - \beta_0, X_i)^\intercal. \tag{A6}$$

Substitution of $\hat{x}_i$ into (A4) gives

$$\frac{(Y_i - \beta_0 - \beta_1 X_i)^2}{(\sigma_{ee} - 2\beta_1\sigma_{eu} + \beta_1^2\sigma_{uu})} \tag{A7}$$

so that after summing over all N points we obtain

$$S = \frac{\sum_{i=1}^N (Y_i - \beta_0 - \beta_1 X_i)^2}{(\sigma_{ee} - 2\beta_1\sigma_{eu} + \beta_1^2\sigma_{uu})}, \tag{A8}$$

which is the expression to be minimized. In (York 1969) the linear equation that minimizes (A8) is given by

$$\beta_1 = \frac{\sum_{i=1}^N Z_i^2 V_i \left[\frac{U_i}{\omega(Y_i)} + \frac{\beta_1 V_i}{\omega(X_i)} - \frac{\gamma_i V_i}{\alpha_i}\right]}{\sum_{i=1}^N W_i^2 U_i \left[\frac{U_i}{\omega(Y_i)} + \frac{\beta_1 V_i}{\omega(X_i)} - \frac{\beta_1 \gamma_i U_i}{\alpha_i}\right]} \tag{A9}$$

where

$$\alpha_i^2 = \omega(X_i)\omega(Y_i), \quad U_i = X_i - \bar{X}, \quad V_i = Y_i - \bar{Y},$$
$$\bar{X} = \left(\sum_{i=1}^N Z_i X_i\right)/\sum_{i=1}^N X_i \quad \text{and} \quad \bar{Y} = \left(\sum_{i=1}^N Z_i Y_i\right)/\sum_{i=1}^N Y_i,$$
$$Z_i = \frac{\omega(X_i)\omega(Y_i)}{\omega(X_i) + b^2\omega(Y_i) - 2b\gamma_i\alpha_i}.$$

Clearly (A9) requires an iterative process to find β_1 which begins with an initial guess which may be found from performing a standard linear regression assuming the X_i to be exact. After β_1 is obtained the value of β_0 is found from the fact that the mean must be on the best fit line and thus $\beta_0 = \bar{Y} - \beta_1\bar{X}$.

The uncertainties in the parameter values, σ_{β_0} and σ_{β_1}, are (Titterington & Halliday 1979)

$$\sigma_{\beta_0}^2 = \frac{\sum Z_i x_i^2}{\left(\sum Z_i x_i^2\right)\left(\sum Z_i\right) - \left(\sum Z_i x_i\right)^2}, \tag{A10}$$

$$\sigma_{\beta_1}^2 = \frac{\sum Z_i}{\left(\sum Z_i x_i^2\right)\left(\sum Z_i\right) - \left(\sum Z_i x_i\right)^2}. \tag{A11}$$

We now return to the transformed model from (12) which takes on the form

$$\log(\delta H_i) = p\log(\langle r_i^2\rangle) + \log\left(\frac{v^2}{2\bar{H}_0}\right) \tag{A12}$$

such that we may identify $y_i = \log(\delta H_i)$, $x_i = \log(\langle r_i^2\rangle)$, and $(\beta_1, \beta_0) = \left(p, \log\left(\frac{v^2}{2\bar{H}_0}\right)\right) = (p, \log A)$.

Since our distances are given in units h^{-1}Mpc, independent of the overall normalization of $\bar{H}_0$, rather than directly performing a fit to (15), (A12) we instead perform a fit to the relation

$$\bar{H}_0\left(H_s' - H_s\right) \approx \frac{v^2}{2}\left(\langle r_i^2\rangle_s\right)^p. \tag{A13}$$

This gives a direct estimate of v in km sec^{-1}, from which we obtain $A \simeq v^2/(2\bar{H}_0)$.