MNRAS 450, 1868–1873 (2015) doi:10.1093/mnras/stv651

An unbiased estimator of peculiar velocity with Gaussian distributed errors for precision cosmology

Richard Watkins[1]⋆ and Hume A. Feldman[2]⋆

[1]Department of Physics, Willamette University, Salem, OR 97301, USA
[2]Department of Physics & Astronomy, University of Kansas, Lawrence, KS 66045, USA

Accepted 2015 March 23. Received 2015 February 28; in original form 2014 November 21

ABSTRACT

We introduce a new estimator of the peculiar velocity of a galaxy or group of galaxies from redshift and distance estimates. This estimator results in peculiar velocity estimates which are statistically unbiased and have Gaussian distributed errors, thus complying with the assumptions of analyses that rely on individual peculiar velocities. We apply this estimator to the SFI++ and the Cosmicflows-2 catalogues of galaxy distances and, since peculiar velocity estimates of distant galaxies are error dominated, examine their error distributions. The adoption of the new estimator significantly improves the accuracy and validity of studies of the large-scale peculiar velocity field that assume Gaussian distributed velocity errors and eliminates potential systematic biases, thus helping to bring peculiar velocity analysis into the era of precision cosmology. In addition, our method of examining the distribution of velocity errors should provide a useful check of the statistics of large peculiar velocity catalogues, particularly those that are compiled out of data from multiple sources.

Key words: galaxies: kinematics and dynamics – galaxies: statistics – cosmology: observations – cosmology: theory – distance scale – large-scale structure of Universe.

1 INTRODUCTION

The Doppler effect provides a remarkably accurate method to infer the velocity of a galaxy towards or away from us by measuring the blue or red-shift of its spectral lines, respectively. However, since cosmological expansion also causes a redshift, determination of the peculiar (local) motion v also requires the measurement of the galaxy's distance r, so that

$$v = cz - H_0 r\,, \tag{1}$$

where c is the speed of light, z is the redshift and H_0 is Hubble's constant. This formula assumes a linear Hubble relation. For more accuracy, particularly at large distances, we can include the effects of cosmic acceleration by replacing z with $z_{\rm mod}$, where

$$z_{\rm mod} = z[1 + 0.5(1 - q_o)z - (1/6)(1 - q_o - 3q_o^2 + 1)z^2]\,, \tag{2}$$

where q_o is the deceleration parameter (see also Davis & Scrimgeour 2014; Springob et al. 2014). In addition, we can achieve additional accuracy by accounting for the fact that redshift is not an additive quantity. Rather than $cz_{\rm mod} = H_0 r + v$, we instead should write $(1 + z_{\rm mod}) = (1 + H_0 r/c)(1 + v/c)$, which reduces to the familiar formula at low redshift. Solving for v, we obtain

$$v = \frac{cz_{\rm mod} - H_0 r}{1 + H_0 r/c} \approx \frac{cz_{\rm mod} - H_0 r}{1 + z_{\rm mod}}, \tag{3}$$

where in the second expression we replaced $H_0 r/c$ with $z_{\rm mod}$, a good approximation since the difference between them, which is approximately v/c, is always much less than 1. The second expression is easier to work with in practice since it does not introduce new factors of r, a quantity that has large uncertainties.

Whereas redshift can be measured very accurately, distance measurements typically have uncertainties of $\simeq$20 per cent, so that the uncertainty, δv, in a peculiar velocity is approximately $\delta v \approx 0.20 H_0 r$. Since typical peculiar velocities are thought to be $\simeq$500 km s^{-1}, we see that for $H_0 \approx 70$ km s^{-1} Mpc^{-1} the uncertainties in peculiar velocities become of the order of their magnitudes for objects at distances $r \gtrsim 35$ Mpc, which includes the region that we would like to use peculiar velocities as a tool to probe large-scale structure. Thus individual peculiar velocity measurements have very low signal-to-noise, which makes it necessary to have a large sample in order to extract meaningful information.

Two major approaches have been used to analyse peculiar velocity catalogues (for reviews, see Jacoby et al. 1992; Strauss & Willick 1995). The first forgoes calculating individual peculiar velocities altogether, and instead uses distance and redshift information to estimate parameters of a model of the peculiar velocity field (for

⋆ E-mail: rwatkins@willamette.edu (RW); feldman@ku.edu (HAF)

recent usage of this method, see Nusser & Davis 2011; Courtois et al. 2012; Hong et al. 2014; Johnson et al. 2014). This approach has the disadvantage that it is difficult to quantify how the different parts of the sample volume are contributing to the final parameter estimates. For example, peculiar velocity samples typically have much more information at small distances than in the outer parts of the sample. This is due both to the higher density of nearby objects and the higher accuracy with which their peculiar velocities can be determined. Thus the results of these analyses can end up mostly reflecting the nearby peculiar velocity field.

The second approach involves combining many individual peculiar velocities into moments of the peculiar velocity field that have much smaller uncertainties, for example, the bulk flow of a sample volume (Kaiser 1988; Lynden-Bell et al. 1988; Courteau et al. 1993, 2000; Feldman & Watkins 1994, 2008; Watkins & Feldman 1995, 2007; Willick 1999; Juszkiewicz et al. 2000; Nusser et al. 2001; Hudson 2003; Feldman et al. 2003; Hudson et al. 2004; Sarkar, Feldman & Watkins 2007). This approach has the advantage that the contribution of different parts of the sample to a moment can be controlled by using various weighting schemes. For example, the Minimum Variance method (Watkins, Feldman & Hudson 2009; Feldman, Watkins & Hudson 2010; Agarwal, Feldman & Watkins 2012; Watkins & Feldman 2015) can be used to create moments that probe a volume in a known, standardized way. It is also straightforward to quantify how these moments probe the power spectrum, making it possible to compare their values to what would be expected, given a particular cosmological model.

This second approach relies on reducing the errors by averaging over many noisy measurements. It is important to emphasize that if the distribution of measurement errors is not symmetric about zero, then the averaging process will result in an incomplete cancellation of the noise leading to a systematic bias, which can suggest an appearance of flows that do not exist, or mask flows that do. In addition, most analyses make the stronger assumption that the error distribution is Gaussian. If this assumption is violated, then the validity of the results could be called into question.

As peculiar velocity surveys become larger and the uncertainties in derived quantities like the bulk flow become smaller, it is important to revisit the validity of our assumptions about the measurement error distribution. For example, peculiar velocities calculated using the traditional estimator are known to have non-Gaussian errors. This comes about because the quantity estimated in distance determinations is actually the distance modulus, the difference between the apparent magnitude and the absolute magnitude, which is related to the logarithm of the distance. The absolute magnitude is determined through its empirical relation to some distance-independent quantity. For example, in the Tully–Fisher (TF) relation, absolute magnitude is related to rotational velocity, or equivalently, emission line widths. The scatter about the empirical relations for absolute magnitude are typically Gaussian, as one would expect from the Central Limit Theorem, leading to Gaussian errors in distance moduli (for the TF relation, see e.g. Masters et al. 2006; Springob et al. 2007). Exponentiating distance moduli to obtain distances skews the distribution of the errors, hence leading also to skewed errors in the peculiar velocities (this issue is also discussed in Johnson et al. 2014). Corrections for homogeneous and inhomogeneous Malmquist bias typically account for this skewness (see e.g. Lynden-Bell et al. 1988; Freudling et al. 1999); however, as we shall see below, in the surveys we examine this correction is not as effective as the new estimator at eliminating skewed tails in the distribution. Skewed, non-Gaussian errors invalidate the statistical assumptions of the analysis methods, and further, may lead to biases in the analyses' results and conclusions.

In this paper, we introduce an unbiased estimator of peculiar velocity that has Gaussian distributed errors. The use of this estimator will greatly increase the accuracy and reliability of any analysis that relies on individual peculiar velocity measurements. We also examine the statistics of several large-scale peculiar velocity surveys with both our new estimator and the traditional estimator to determine the validity of our assumptions about measurement errors.

In Section 2, we describe in detail the peculiar velocity estimator. In Section 3, we discuss the statistics of peculiar velocity surveys. We conclude in Section 4.

2 PECULIAR VELOCITY ESTIMATOR

Our goal is to obtain an estimate, v_e, of the peculiar velocity of a galaxy or group from the galaxy's redshift cz and an estimate of its distance r_e. Given equation (1), the most straightforward estimator is

$$v_e = cz - H_0 r_e \,, \tag{4}$$

and this is typically the estimator used in peculiar velocity analyses. However, from a statistical point of view, this estimator has several undesirable qualities. (For a general discussion of the statistics of estimators, see Lupton 1993.) First, distance indicators give distance moduli or log-distances with Gaussian distributed errors. Exponentiating skews the error distribution, resulting in distance errors that are not Gaussian distributed. Secondly, this estimator is biased in a statistical sense: the average of an ensemble of velocity estimates with different errors is not the true value, i.e. $\langle v_e \rangle \neq v$. This is the result of the skewness of the distribution of distance errors, which gives rise to $\langle r_e \rangle \neq r$. These undesirable features can lead to biases in our analyses and in general invalidate our statistical assumptions about the errors in peculiar velocities. They suggest that we should be investigating other estimators that might be better behaved statistically.

Instead we propose calculating peculiar velocities using the estimator

$$v_e = cz \log(cz/H_0 r_e) \,. \tag{5}$$

While this estimator may look unfamiliar, it has the statistical properties that we desire in an estimator. First, since it uses the log-distance (or equivalently, the distance modulus), it has Gaussian distributed errors. It is easy to see that the uncertainty in the peculiar velocity, δv_e, is given by $\delta v_e = cz\delta l_e$, where δl_e is the uncertainty in the log-distance. Secondly, we can use $\langle \log(r_e) \rangle = \langle \log r \rangle$ to show that this estimator is unbiased, as long as the true $v \ll cz$, which is a good assumption for distant galaxies,

$$\begin{aligned}
\langle v_e \rangle &= -cz(\langle \log(H_0 r_e) \rangle - \log(cz)) \\
&= -cz(\log(H_0 r) - \log(cz)) \\
&= -cz(\log(cz - v) - \log(cz)) \\
&= -cz(\log(1 - v/cz)) \\
&\approx v \,,
\end{aligned} \tag{6}$$

where we have used equation (1) to replace $H_0 r$ with $cz - v$, and we have assumed that the uncertainties in the redshift cz are negligible.

From equation (3), we see that a more accurate estimator at large redshift is given by

$$v_e = \frac{cz_{\rm mod}}{(1+z_{\rm mod})}\log(cz_{\rm mod}/H_0 r_e)\,, \tag{7}$$

with uncertainty $\delta v_e = cz_{\rm mod}\delta l_e/(1+z_{\rm mod})$. We stress that the assumption that we are making is that the *actual* velocity of the galaxy or group (v) is small compared to the redshift, not the estimated velocity (v_e). While estimates of peculiar velocities can be a few$\times 10^3$ km s^{-1}, it is thought that most actual peculiar velocities are at most a few$\times$ 10^2 km s^{-1}. Our assumption should hold quite well for galaxies at distances $\gtrsim$ 20 Mpc.

3 STATISTICS OF PECULIAR VELOCITY SURVEYS

The defining characteristic of large-scale peculiar velocity surveys is that they have low signal-to-noise ratio. However, if our goal is to determine the distribution of the noise, then we can turn this to our advantage. In particular, if we consider objects with peculiar velocity errors σ such that $\sigma \gg \sigma_v$, where σ_v is the spread in actual peculiar velocities, then we can be assured that the objects' measured peculiar velocities are dominated by noise, with negligible contributions from actual motions.

Here we will examine the error distribution in two large peculiar velocity surveys. The SFI++ (Masters et al. 2006; Springob et al. 2007) is a sample of 4052 spiral galaxies with TF distances. The Cosmicflows2 (hereafter CF2; Tully et al. 2013) galaxy catalogue is a compendium of distances to 8135 galaxies measured with various methods, including TF, Fundamental Plane, SNIa, surface brightness fluctuations and tip of the red giant branch (TRGB). While the CF2 contains the SFI++ as one of its largest components, in compiling the CF2 a reanalysis of the literature distances was done to ensure consistency between data sets. For both samples, we use the more accurate expressions given by equation (3) for the old estimator and equation (7) for the new estimator, following Tully et al. (2013) in assuming the standard cosmological model with $\Omega_m = 0.27$ and $\Omega_\Lambda = 0.73$, so that $q_o = 0.5(\Omega_m - 2\Omega_\Lambda) = -0.595$.

Another difference in the catalogues is that the SFI++ catalogue provides distances in km s^{-1} and so are scaled relative to the Hubble constant. In contrast, the CF2 sample attempts to determine an absolute scale, and so reports distances in Mpc. Thus to calculate peculiar velocities from the distances in the CF2, we must assume a value for the Hubble constant. The nominal value given by the authors of the CF2 in Tully et al. (2013) is 74.4 km s^{-1} Mpc^{-1}.

In Fig. 1, we show histograms for the values of the peculiar velocity divided by their uncertainty, v_i/σ_i, calculated using both the new estimator and the traditional estimator for galaxies with $\sigma_i >$ 1000 km s^{-1} in the CF2 survey. If our statistical assumptions are correct, and if actual motions make only a small contribution, then the values v_i/σ_i should be unit Gaussian variates, and the histograms should match the Gaussian of unit standard deviation shown in the figure. We see that the histogram using the new estimator is a good match to the unit Gaussian, but that the traditional estimator results in a skewed distribution.

As seen in the figure, the exponentiation of the Gaussian distributed log-distances results in a distribution of errors that is skewed in a complicated way, with the peak shifted towards negative velocities but with a shortened tail on the negative side and an elongated tail on the positive side. This effect is more clearly seen in Fig. 2 where we plot the histogram using a logarithmic scale. This skewness cannot be corrected for by simply shifting the centre of the distribution. Nor can the skewness be corrected by adjusting only negative velocities, as is proposed by Tully et al. (2013) to correct for what they call 'error bias'.

In Fig. 3, we show v_i/σ_i histograms for galaxies in the SFI++ survey, again for $\sigma_i >$ 1000 km s^{-1}. As in Fig. 2, we plot the histograms on a logarithmic scale to accentuate the tails of the distributions. The catalogue provides both Malmquist-corrected and -uncorrected distances. Malmquist bias correction methods account for the skewness of the distribution of distance errors, and Lynden-Bell et al. (1988) showed that velocities calculated with Malmquist-corrected distances should be approximately Gaussian distributed. It does not make sense to use Malmquist-corrected distances with the new estimator, since, as we discuss below, different corrections apply for distance moduli or log-distances than for distances. We

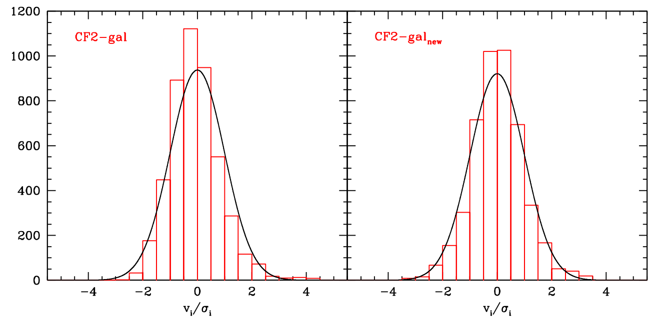

Figure 1. The histograms for the values of the peculiar velocity over their uncertainty, v_i/σ_i, calculated using both the traditional estimator (left-hand panel) and the new estimator (right-hand panel) for galaxies with $\sigma_i >$ 1000 km s^{-1} in the CF2 survey. The area of the histograms is normalized to unity. We also show a Gaussian of unit area and unit variance.

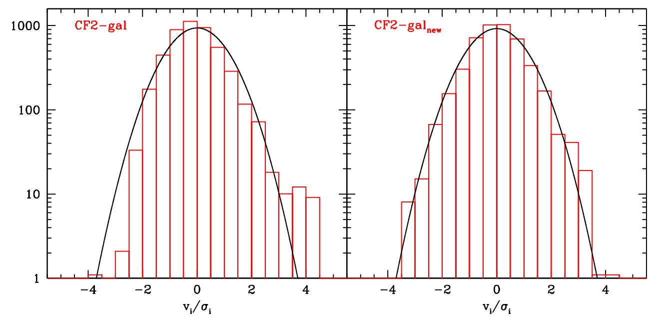

Figure 2. Same as Fig. 1 using the logarithm of the histograms to show better the behaviour of the tails of the Gaussian distribution of the traditional (left-hand panel) and new (right-hand panel) estimators.

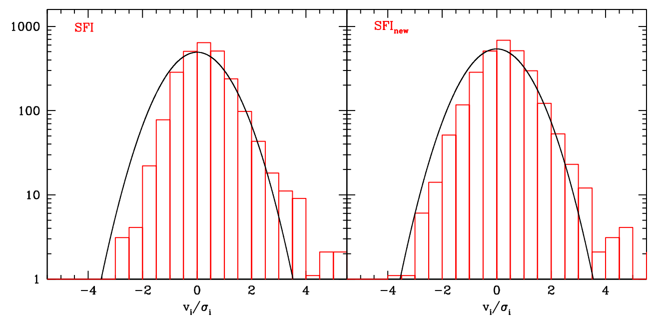

Figure 3. The same as Fig. 2 for the SFI++ survey. We show both the Malmquist-corrected traditional estimator (left-hand panel) and Malmquist-uncorrected new estimator (right-hand panel) distances.

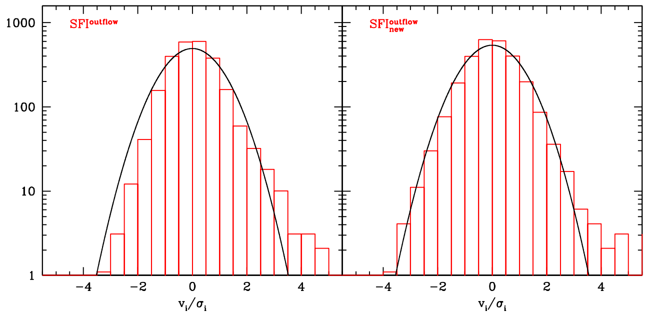

Figure 4. The same as Fig. 3 with a 400 km s^{-1} coherent outflow in the survey volume added.

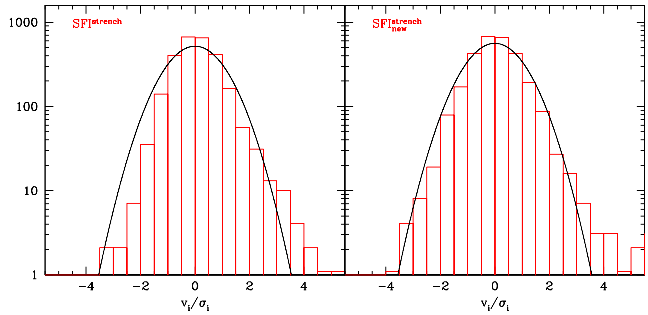

Figure 5. The same as Fig. 3 with all distances increased by 5 per cent for the SFI++ survey.

thus show histograms for the new estimator using uncorrected values and the traditional estimator using corrected values. First, we see that the histograms for the new estimator and the traditional estimator using Malmquist-corrected distances do in fact have the same peak. However, the histograms are not centred on zero.

One possible explanation for the skewing of the error histograms from zero in Fig. 2 is a coherent outflow in the volume occupied by the survey. While random velocities, or even bulk motions, should have little effect on the histograms since their affect would average out over different directions, a coherent outflow would be expected to shift the peak of the histograms towards positive velocities, exactly the effect we see in the figure. We can test this explanation by considering a simple model where the outflow adds a constant peculiar velocity to each galaxy in the survey. In Fig. 4, we show the same histograms as in Fig. 2 except that we have subtracted 400 km s^{-1} from the peculiar velocity of each galaxy. We see that subtracting a relatively modest outflow has resulted in an error distribution that is Gaussian and centred on zero. The existence of a coherent outflow would support recent arguments that we live in a low-density region, dubbed the 'local hole' (Whitbourn & Shanks 2014).

The disagreement between the CF2 and SFI++ catalogues regarding the existence of a coherent outflow is a consequence of the addition of new data and the reanalysis of literature distances that was done when the CF2 catalogue was assembled. Tully et al. (2013) compared distance estimates for galaxies and groups that appeared in more than one component sample and used this comparison to rescale and reanalyze distances to achieve statistical consistency between all the components of the CF2 sample. This reanalysis was anchored by the zero-point provided by Cepheid and TRGB distances. They found that the resulting CF2 catalogue was consistent with a Hubble constant $H_0 = 74.4$ km s^{-1} Mpc^{-1} that did not vary with redshift. It is worth noting that although this relatively low redshift ($cz \lesssim 0.1$) measurement of H_0 is in tension with microwave background results, it agrees well with a recent measurement of H_0 using SNIa at much higher redshift (Neill et al. 2014).

It is possible that the rescaling of the SFI++ could have inadvertently 'erased' a real coherent outflow. However, this suggests another explanation of the skewness in the error distribution of the SFI++ survey, a systematic error in the scaling of distances. In Fig. 5, we show the same histograms as in Fig. 3, but with all distances increased by 5 per cent. This scaling is equivalent to changing the zero-point of the SFI++ or increasing its Hubble constant. Again, we see that there is good agreement between the new estimator histogram and the unit Gaussian centred on zero. This roughly corresponds to the rescaling of the SFI++ in the CF2. A direct comparison of the distances given in the SFI++ and the distances for the same galaxies in the CF2, using $H_0 = 74.4$ km s^{-1} Mpc^{-1}, shows that CF2 distances are about 6.8 per cent larger on average.

In both Figs 4 and 5, we see that although the histogram using the traditional estimator with Malmquist-corrected data is indeed approximately Gaussian, the tails of this distribution are still noticeably skewed. This demonstrates that our new estimator is more effective at correcting for the skewness of peculiar velocity errors than the Malmquist bias corrections used in the SFI++ survey. Since these correction methods are substantially similar to those used in other surveys, it is likely that this is true in general.

4 DISCUSSION

Peculiar velocity analysis methods that work with velocity measurements for individual galaxies, groups or clusters assume that the errors in velocity measurements have a Gaussian distribution. However, the estimator that is traditionally used is known to have a skewed, non-Gaussian error distribution. Malmquist bias corrections include a correction that shifts the peak of the error distribution to zero, but these corrections do not remove the skewness in the tails of the distribution. These tails are particularly important since they represent objects with measurement errors that are larger than expected given their uncertainties. Given that measurement uncertainties typically dominate over the true velocities, these objects have large ratio of estimated peculiar velocity to uncertainty. Since analyses of velocity moments typically weight by uncertainty, these velocities may have a large impact on results, and can potentially introduce biases if the uncertainties are not distributed symmetrically about the central value.

As peculiar velocity catalogues become larger, with a corresponding decrease in the calculated uncertainties in low-order moments such as the bulk flow, it becomes increasingly important to address potential systematic errors arising from non-Gaussian velocity error distributions. We have introduced a simple, easy-to-use peculiar velocity estimator that results in velocities with unbiased, Gaussian errors. We have shown that this estimator works well when applied to the CF2 catalogue and, with some adjustment, the SFI++ catalogue of galaxy distances. This new estimator is an important step in bringing peculiar velocity analyses into the era of precision cosmology.

Our new estimator should not be used to estimate peculiar velocities with Malmquist-corrected distances, since currently implemented Malmquist correction procedures already account for the skewness of the traditional estimator. While we have shown that Malmquist correction does result in approximately Gaussian velocity errors in the SFI++ survey, we have seen that our new estimator does a better job of producing a distribution with symmetric tails.

Modifying Malmquist correction methods to be used with the new estimator is straightforward. Specifically, in the SFI++ survey, Malmquist bias corrections to distances are implemented by calculating the corrected probability $p(r_i)$ of a galaxy being at a distance r_i through the convolution (Springob et al. 2007)

$$p(r_i) = k_1 p_{\rm TF}(r_i) p_{\rm mag}(r_i) p_{\rm lss}(r_i), \tag{8}$$

where k_1 is a normalization constant, $p_{\rm TF}(r_i)$ is the probability distribution for the galaxy being at a position r_i as given by the TF measurement, $p_{\rm mag}(r_i)$ is the probability of finding a galaxy with its apparent magnitude at a distance r_i and $p_{\rm lss}(r_i)$ is the density distribution along the line of sight as given by a redshift survey. To use the new peculiar velocity estimator, we instead calculate the corrected probability $p(\mu_i)$ of finding a galaxy with a distance modulus μ_i as

$$p(\mu_i) = k_1 p_{\rm TF}(\mu_i) p_{\rm mag}(\mu_i) p_{\rm lss}(\mu_i), \tag{9}$$

where the probabilities $p_{\rm TF}$, $p_{\rm mag}$ and $p_{\rm lss}$ have now been expressed in terms of the distance modulus. Note that in these expressions $p_{\rm TF}(\mu_i)$ is calculated using the distance modulus, while $p_{\rm TF}(r_i)$ is not. Note also that the maximum of $p(r_i)$ will not correspond to the distance modulus that maximizes $p(\mu_i)$, so that Malmquist-corrected values of r_i cannot be used in our new estimator; instead, Malmquist-corrected values of μ_i must be calculated using $p(\mu_i)$.

As a specific example, consider the simple case of a uniform density of galaxies, where, for mathematical simplicity, we will work with the equivalent log-distance $l_i = \log(r_i)$ instead of the distance modulus μ_i. In this case there is a bias, sometimes called the *homogeneous Malmquist bias*, whereby galaxies are more likely to have scattered from larger than smaller radius due the increasing This comes into our calculations through the fact that in this case $p_{\rm lss}(r_i){\rm d}r_i \propto r^2{\rm d}r_i$, so that $p_{\rm lss}(l_i){\rm d}l_i \propto {\rm e}^{3l_i}\,{\rm d}l_i$. Assuming that $p_{\rm mag}(l_i)$ is constant and that $p_{\rm TF}(l_i)$ is a Gaussian distribution centred on the value l_o with uncertainty Δ, we have

$$\begin{aligned} p(l_i) &\propto \exp(-(l_i - l_o)^2/2\Delta^2){\rm e}^{3l_i} \\ &\propto \exp(-(l_i - (l_o + 3\Delta^2))^2/2\Delta^2). \end{aligned} \tag{10}$$

Thus we see that $p(l_i)$ remains Gaussian, with the effect of the Malmquist bias correction being to shift the peak of the distribution outwards by $3\Delta^2$. The size of this shift matches the result of a similar calculation given in Lynden-Bell et al. (1988). More generally, this calculation suggests that as long as the product $p_{\rm mag}(r_i)p_{\rm lss}(r_i)$ can be approximated by a power law in r_i in the region around the galaxy's location, the effect will be to shift the peak of $p(\mu_i)$ relative to $p_{\rm TF}(\mu_i)$ while maintaining a Gaussian distribution. Since $p_{\rm mag}$ and $p_{\rm lss}$ are typically slowly varying compared to $p_{\rm TF}$, it is thus reasonable to expect that Malmquist-corrected μ_i will still have Gaussian errors. We will investigate this issue in more detail in future work.

Large-scale motion analyses require estimates of both the radial peculiar velocities and positions of a set of galaxies. While we must use both redshift and a distance estimate to calculate peculiar velocity, either of these quantities can be used to estimate position. Analyses that use distance estimates to estimate position are said to be done in 'real space', while those that use redshift are said to be done in 'redshift space'. While it may seem more intuitive to use a distance estimate to estimate position, it is important to remember that redshift is often a more accurate estimate of distance, particularly in situations where distance units of h^{-1} Mpc or km s^{-1} are used, so that uncertainty in the value of the Hubble constant does not enter the calculation. Due to the small uncertainty in redshift, the scatter of the redshift about the distance (as measured in km s^{-1}) is caused almost entirely by peculiar velocities, which are thought to be of the order of 500 km s^{-1} at most. This is to be contrasted with distance estimates, which for many distance indicators have uncertainties of the order of 20 per cent. Thus redshift begins to be a more accurate measure of a galaxy's position around distances of 2500 km s^{-1}, or 25h^{-1} Mpc, and for distances of the order of 100 h^{-1} Mpc, redshift can be a factor of 4 more accurate than distance estimates on average. We note that the advantage of redshift over distance estimates can be somewhat smaller for clusters and groups of galaxies, where distance uncertainties can be reduced by $\sqrt{N}$, where N is the number of galaxies in the cluster with measured distances, and for SNIa, where distance uncertainties are closer to 5 per cent. Because position uncertainties are much smaller in redshift space, particularly for objects at large distances, Malmquist bias effects that are caused by position uncertainties are much less important and can be neglected. However, in redshift space one must account for other forms of Malmquist bias that affect the determination of distance relation parameters, e.g. the slope and zero-point of the TF relation. For a more detailed discussion of Malmquist bias in real and redshift space, see Strauss & Willick (1995).

The new estimator we introduced here will prove particularly useful for peculiar velocity analyses that are done in redshift space with data that has not been Malmquist corrected. In this case, peculiar velocities calculated with the traditional estimator have an error distribution that is biased in addition to being skewed. For example, the estimator we have introduced alleviates the problem of 'error bias' noted in Tully et al. (2013). We have shown that peculiar velocities calculated from CF2 distances using the new estimator have a symmetric, Gaussian error distribution and do not require any further correction.

We have also presented a method to check large catalogues of peculiar velocities to confirm that they have the expected distribution of errors. We stress that skewness or non-Gaussianity in velocity error distributions can lead to results which do not accurately reflect the large-scale flows we are trying to study. This method should provide a useful tool for compiling large peculiar velocity catalogues, particularly when combining data from different sources.

Finally, we have seen that the distribution of errors in the SFI++ survey is not centred on zero. This can be explained by an approximately 400 km s^{-1} coherent outflow in the survey volume or by a systematic error in the scaling of distances of about 5 per cent. Which of these explanations is correct is an interesting question that should be pursued in further research.

ACKNOWLEDGEMENTS

We would like to thank Brent Tully and Mike Hudson for useful comments.

REFERENCES

Agarwal S., Feldman H. A., Watkins R., 2012, MNRAS, 424, 2667
Courteau S., Faber S. M., Dressler A., Willick J. A., 1993, ApJ, 412, L51
Courteau S., Willick J. A., Strauss M. A., Schlegel D., Postman M., 2000, ApJ, 544, 636
Courtois H. M., Hoffman Y., Tully R. B., Gottlöber S., 2012, ApJ, 744, 43

Davis T. M., Scrimgeour M. I., 2014, MNRAS, 442, 1117
Feldman H. A., Watkins R., 1994, ApJ, 430, L17
Feldman H. A., Watkins R., 2008, MNRAS, 387, 825
Feldman H. et al., 2003, ApJ, 596, L131
Feldman H. A., Watkins R., Hudson M. J., 2010, MNRAS, 407, 2328
Freudling W. et al., 1999, ApJ, 523, 1
Hong T. et al., 2014, MNRAS, 445, 402
Hudson M. J., 2003, preprint (arXiv:astro-ph/0311072)
Hudson M. J., Smith R. J., Lucey J. R., Branchini E., 2004, MNRAS, 352, 61
Jacoby G. H. et al., 1992, PASP, 104, 599
Johnson A. et al., 2014, MNRAS, 444, 3926
Juszkiewicz R., Ferreira P. G., Feldman H. A., Jaffe A. H., Davis M., 2000, Science, 287, 109
Kaiser N., 1988, MNRAS, 231, 149
Lupton R., 1993, Statistics in Theory and Practice. Princeton Univ. Press, Princeton, NJ
Lynden-Bell D., Faber S. M., Burstein D., Davies R. L., Dressler A., Terlevich R. J., Wegner G., 1988, ApJ, 326, 19
Masters K. L., Springob C. M., Haynes M. P., Giovanelli R., 2006, ApJ, 653, 861
Neill J. D., Seibert M., Tully R. B., Courtois H., Sorce J. G., Jarrett T. H., Scowcroft V., Masci F. J., 2014, ApJ, 792, 129
Nusser A., Davis M., 2011, ApJ, 736, 93
Nusser A., da Costa L. N., Branchini E., Bernardi M., Alonso M. V., Wegner G., Willmer C. N. A., Pellegrini P. S., 2001, MNRAS, 320, L21
Sarkar D., Feldman H. A., Watkins R., 2007, MNRAS, 375, 691
Springob C. M., Masters K. L., Haynes M. P., Giovanelli R., Marinoni C., 2007, ApJS, 172, 599
Springob C. M. et al., 2014, MNRAS, 445, 2677
Strauss M. A., Willick J. A., 1995, Phys. Rep., 261, 271
Tully R. B. et al., 2013, AJ, 146, 86
Watkins R., Feldman H. A., 1995, ApJ, 453, L73
Watkins R., Feldman H. A., 2007, MNRAS, 379, 343
Watkins R., Feldman H. A., 2015, MNRAS, 447, 132
Watkins R., Feldman H. A., Hudson M. J., 2009, MNRAS, 392, 743
Whitbourn J. R., Shanks T., 2014, MNRAS, 437, 2146
Willick J. A., 1999, ApJ, 516, 47

This paper has been typeset from a TEX/LATEX file prepared by the author.