Perturbation Analysis

Christopher Rooney

4 March 2016

ABSTRACT

We have explored several different methods of perturbing the distances (and therefore velocities) of galaxies in order to best match the perturbation effects that impact real surveys. Here I will report my progress, including the results from several different perturbation methods and detail some of the problems with each of them.

1 INTRODUCTION

I will be using several different methods of perturbing a galaxy's distance, and subsequently several different methods of calculating the new velocity of that galaxy. To keep things clear, I will give each a concise name and implementation here.

Three different types of distance modulus are in use: the **log** method $\mu_{log} = \log(d)$, the **ln** method $\mu_{ln} = \ln(d)$, and the **actual** method $\mu_a = 5\log(d) + 25$. These names don't necessarily reflect the status of any given formula outside this paper, they are only being used for internal consistency. Also, moduli from the log method can be related to moduli from the ln method by a constant multiple, $\mu_{log} = c \cdot \mu_{ln}$, where $c = log(e)$.

The first method will be referred to as the **distance** method. To perturb a sample using the distance method, a relative distance error is chosen, $\delta d/d$. A normally distributed random number generator is then created with mean d, the galaxy's perfectly known distance (from the simulation), and standard deviation $d \cdot (\delta d/d)$. A value is taken from this RNG and used as the galaxy's distance. This method is only used with the naive estimator, because the unbiased estimator would not make sense here.

The second method is the **relative** method. First, a relative error $\delta\mu/\mu$ is chosen. Then, the distance modulus μ is computed for a galaxy at distance d. A normally distributed RNG is created with mean μ and standard deviation $\mu \cdot (\delta\mu/\mu)$. By this method, galaxies with larger distance moduli will havee much greater spreads of perturbed distances.

The final method is the **modulus** method. This time, instead of picking a relative error, a constant modulus error is chosen, $\delta\mu$. The normal distribution is set up with mean μ and standard deviation $\delta\mu$. A constant error in the modulus should become a relative error in the distance when the distance is computed from the perturbed modulus.

Both the relative and modulus perturbation methods can be used with any one of the distance modulus formulas. However, in order to get consistent results the chosen relative errors have to compensate for the choice of modulus, as detailed in my report "Testing Moduli".

Finally, there are two different ways to convert a perturbed distance and true redshift into a velocity. The **naive** method is the standard formula $cz = h_0 d + v$ solved for velocity. The **unbiased** estimator is detailed in Feldman 2015.

2 DATA

Surveys of $\approx$4000 galaxies are created using the Millennium Simulation data to resemble the CosmicFlows-2 surveys in terms of spatial distribution. A typical galaxy from the Millennium Simulation has a very small peculiar velocity that is greatly exaggerated by the perturbations applied. Fig 1 shows the distribution of peculiar velocities for an unperturbed sample. Fig 1 is on the same scale as all of the other plots in the paper, which is not the best choice for that specific set of galaxies. For this reason, Fig 2 shows the same information on a scale better suited for it, but not the same as all of the other plots.

The other initial sample that was used for this analysis was the COMPOSITE survey of $\approx$ 4000 galaxies. Its characteristics are shown in Fig 5. It is apparent that there are many galaxies with much larger peculiar velocities than are seen in the simulation, due to the real problems that arise when measuring the distance to a galaxy. Fig 5 also shows the peculiar velocity plotted against redshift instead of distance, which can be more accurate at large distances.

Now we present the results from different perturbation methods. The results of the distance method are presented in Fig 6. These plots show similar behavior to Fig 5. I am concerned that the velocities are biased negatively when plotted against redshift. It appears that the dense region of galaxies is spread to more negative velocities at larger distances, which is an effect that we will be seeing a lot. This effect vanishes when the velocity is plotted against redshift.

The negative slope is much more exaggerated in Figs 7 and 8, which show, respectively, the galaxies perturbed by the relative method and the modulus method. Again, the effect vanishes when redshift is used on the x-axis, which indicates that it is an effect caused by some property of the distance calculation, not necessarily the velocity calculation.

At first I wasn't convinced that this negative slope was a
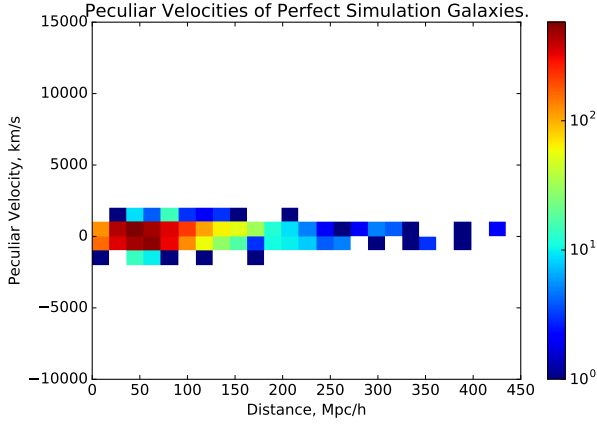
Figure 1. An unperturbed sample of galaxies. The detailed structure of this sample is invisible because the scale is set for consistency with future plots. It is apparent, however, that no galaxies exist in the simulation with a peculiar velocity toward or away from us greater than 2500 km/s. Because of our perfect knowledge of these galaxies, a plot that uses redshift as distance would be exactly the same as this one, with a different x scaling factor.
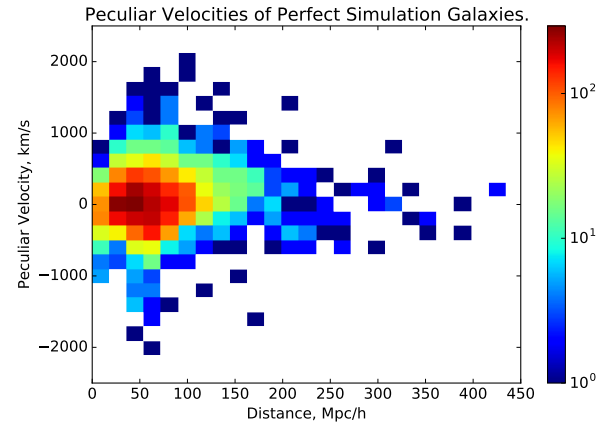

Figure 2. An unperturbed sample of galaxies. This plot does not have the same y-axis scaling as the rest of the plots in order to better show the structure of the unperturbed sample of galaxies.

problem, but I ended up exploring the parameter space a bit more and found that it is definitely a problem. I made two exaggerated plots, one using the distance method and one using the relative method. These are shown in Figs 3 and 4. The bias here is absolutely terrifying.

3   DISCUSSION AND CONCLUSIONS

There are obviously problems with using these methods. If the real sample was perturbed in the modulus, some corrections must have been applied to ensure that the data is not biased. The bias in these methods causes galaxies at larger dis-
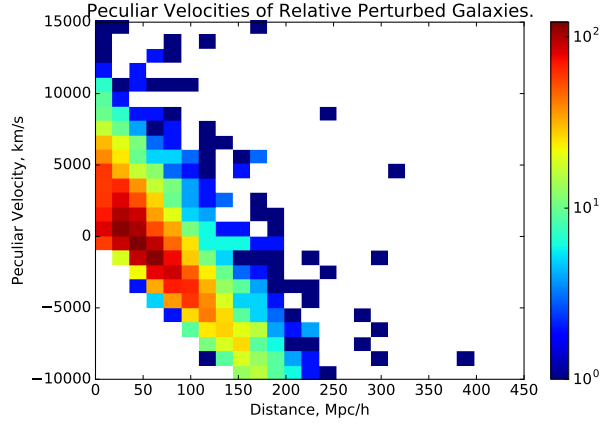

Figure 3. The galaxy sample perturbed using the relative method with an exaggerated constant of 0.2 to show the bias.
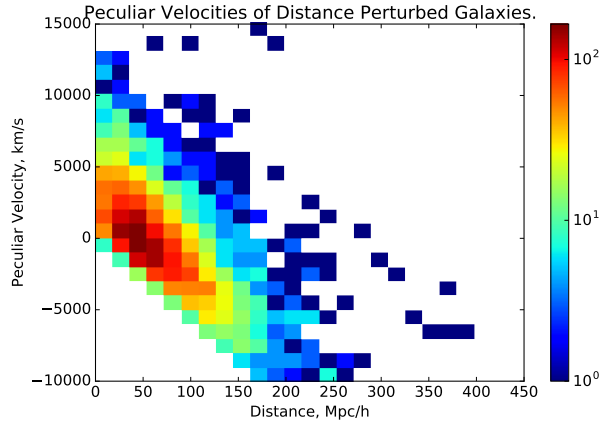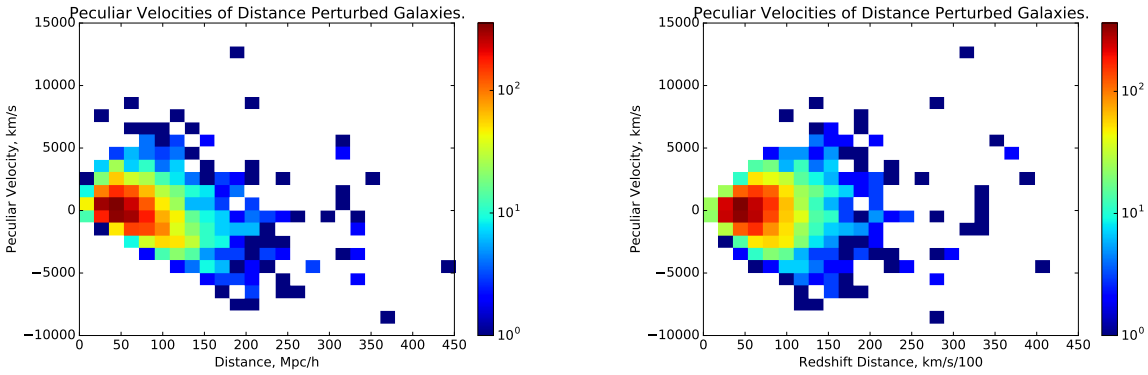

Figure 4. The galaxy sample perturbed using the distance method with an exaggerated constant of 0.5 to show the bias.

tances to appear more blueshifted than is observed, and this deficiency in the perturbation method must be fixed before any meaningful results can be expected from correlations involving them. My current working guess/explanation is based off of Yuyu's idea. The peculiar velocities are estimated using an unbiased estimator, but the distances aren't. Because of that, the distances are skewed strongly, producing this negative slope in the distance plots of the relative and modulus perturbed data.

However, that isn't all because this bias appears in the distance-perturbed sample as well, which never contains a log. I have only a little experience with Malmquist bias, but it's all over the place. Could it be affecting us here?
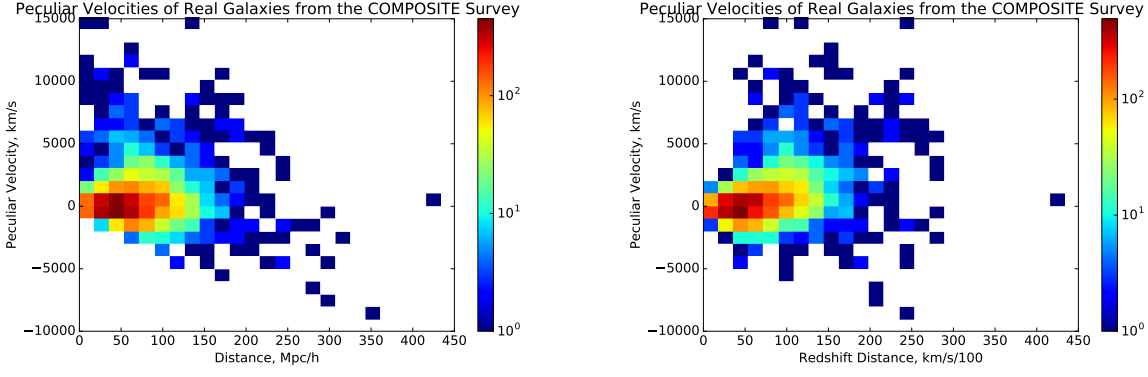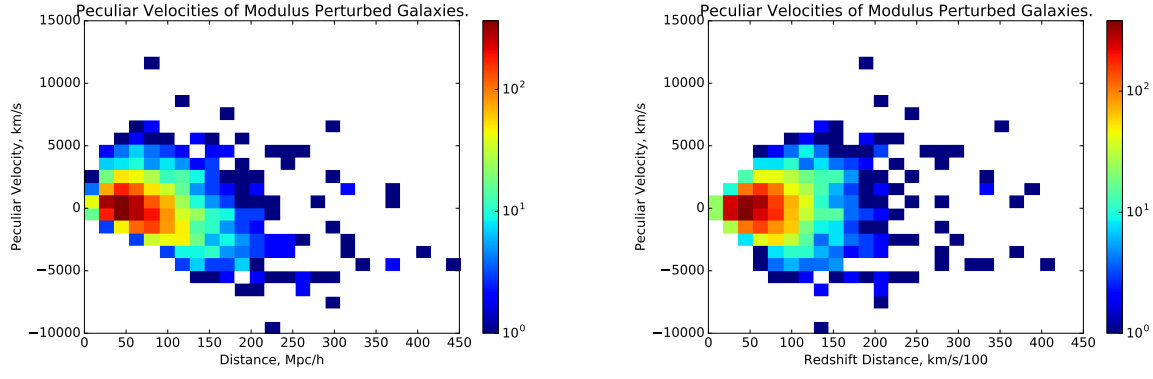
Figure 5. A real sample of galaxies. The spread in velocity is much greater than the simulation, but the simulation and sample both stay centered on zero over large distances. On the right redshift is used instead of distance to present the peculiar velocities of galaxies. At large distances, redshift is a better estimator of distance than distance is, due to the relatively low peculiar velocities of galaxies compared to cosmic expansion.

Figure 6. The perfect sample of galaxies perturbed using the distance method with $\delta d/d = 0.2$, which is approximately the same value used in COMPOSITE. On the left the velocities are plotted against distance, on the right they're plotted using redshift as distance.
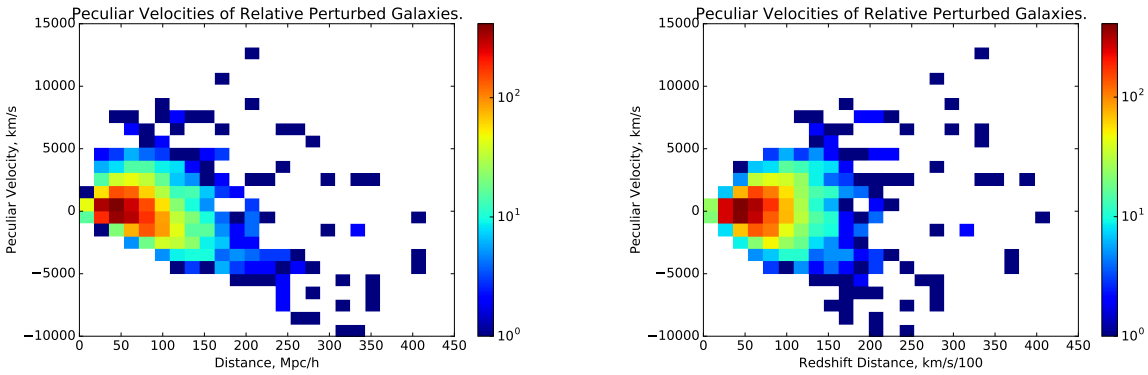
Figure 7. The perfect sample of galaxies perturbed using the relative method with $\delta \mu/\mu = 0.04$, which is the value that ill give the same results as those in the other plots. On the left the velocities are plotted against distance, on the right they're plotted using redshift as distance.

Figure 8. The perfect sample of galaxies perturbed using the modulus method with $\delta d / d = 0.4$, which is the value that will give the same results as those in the other plots. On the left the velocities are plotted against distance, on the right they're plotted using redshift as distance.