Testing Different Methods for Estimating Peculiar Velocities using Simulations

by

Albert Ramon Jr

Submitted in Partial Fulfillment of the
Requirements for the Degree

Bachelor of Science

Supervised by
Dr. Richard Watkins

Department of Physics

Willamette University, College of Liberal Arts
Salem, Oregon

2022

Presentations and publications:

- A. Ramon, *Testing Different Methods for Estimating Peculiar Velocities using Simulations*, Oral Presentation, Willamette University, Spring 2021.

- A. Ramon, *Testing Methods of Estimating Peculiar Velocities from Distance Modulus and Redshift Measurements*, Oral Presentation, Willamette University, Fall 2021.

Acknowledgments

First and foremost, I would like to thank my thesis advisor Richard Watkins who has been so supportive during this thesis endeavor. Professor Watkins has pushed me to heights I did not think I could achieve and I truly appreciate him for being there with me every step of the way.

Thank you to my advisor Michaela Kleinert for also believing in me and helping me throughout this crazy college journey. Your dedication to your students does not go unnoticed and I appreciate all that you did for me.

I would also like to thank the entire Physics department for such a great college experience and all of the help that each and every one of you gave me at some point in time.

Thank you to the National Science Foundation under grant AAG-1907365 for allowing me to do research work towards my thesis beginning in the summer of 2021.

Finally, I would like to thank my parents, Albert and Shannon Ramon. Without you guys none of this would have been possible. Thank you for always being there for me and supporting me in everything I do. No amount of words could express how thankful I am for you two, but just know that I am extremely honored and blessed to call you my parents. I would also like to thank my aunt and uncle, Ron and Kim Timperio. Thank you for always being there for me and always willing to give me support when I needed it the most. Thank you to my brother and sisters, Rob, Melissa, and Michelle for always being there for me when I needed someone to talk to and for constantly supporting me. To all of my family, thank you all for constantly giving me a reason to push through. I did this for you guys!

A big thanks to both the football and track and field programs for giving me a great student-athlete experience. To my dear friends, thank you guys for always being there and for being real ones throughout. I appreciate everyone, foreals!

Abstracts

General Abstract

The measurements of large numbers of galaxy velocities can allow us to study large scale flows through the Universe and test whether they are consistent with the standard Big Bang model. Galaxy velocities are measured using redshifts; however, distances to galaxies must also be measured in order to subtract off the contribution to the redshift due to the expansion of the Universe. In this project we will be using two different methods, one that is a formula and another one that is based upon a probability distribution. We will be comparing the two methods to see which one can better estimate velocity. After conducting a thorough analysis through a series of code, we can conclude that both methods work at similar rates; however, the probability methods works better for smaller redshift distances than the formula method.

Technical Abstract

The measurements of large numbers of galaxy velocities can allow us to study large scale flows through the Universe and test whether they are consistent with the standard Big Bang model. Galaxy velocities are measured using redshifts; however, distances to galaxies, denoted by r, must also be measured in order to subtract off the contribution to the redshift due to the expansion of the Universe. Through the use of the distance modulus μ, we will have non-Gaussian distributed errors, due to the use of an exponential function to convert between μ and r. This then leads to a bias in the estimation of peculiar velocities. This project aims to correct this bias through two different methods to better estimate the peculiar velocity. The first method is based upon a formula that accounts for the bias through the use of logarithmic functions. The second method uses probability distributions to find the peculiar velocity though an updated distance measurement, r, which is unbiased. We will be comparing the two methods to see which one, if any, better correct the distance conversion bias and estimation of peculiar velocity. After conducting a thorough analysis through a series of code, we

can conclude that both methods work at similar rates; however, the probability methods works better for smaller redshift distances than the formula method.

Table of Contents

List of Figures

1 Introduction

For a long time we have grown to accept the idea that we do not live in an extraordinary place and that we are not at the focal point of the Universe. This perspective permits us to make some expansive suppositions. For instance, the Universe looks the same all over. This is true on large scales; however, if we look at the universe on smaller scales there are differences. Truth be told we can take that assertion to a limit and expect that at a large enough scale, the Universe appears to be identical at each and every point in space .Such a space-time is named to be homogeneous. Fig. 1.1 is an image from the 2dF Galaxy Redshift Survey, which said and shows the idea that our universe is homogeneous, as both sides of the graph are nearly identical, signifying homogeneity throughout.

There is another presumption that considers the outrageous homogeneity of the Universe and that is the way that, at some random point in space, the Universe appears to be identical toward whatever path we look. Again such a suspicion can be taken to a limit so that at any time, the Universe appear to be identical, whatever heading one looks. Such a space time is said to be isotropic.

Homogeneity and isotropy are related concepts. For instance a Universe which is isotropic will not be homogeneous while a Universe that is homogeneous is also isotropic. For instance, on the off chance that the Universe is isotropic, this implies you will see no distinction in the structure of the Universe as you glance various ways(at least over large scales). Homogeneity, when seen on the larger scales, implies that whether you look at Universe from Earth or a galaxy a million light years away, it will look the same throughout. Notice that this is clearly not true for the Universe on small scales such as the size of the Earth and even the size of our Galaxy. They look the same only on very large scales. For cosmology, we only consider the isotropy and homogeneity of the Universe on scales of millions of light-years. A universe that is both homogeneous and isotropic is said to fulfill the Cosmological Principle. It is accepted that our Universe fulfills the Cosmological Principle. The Cosmological Principle is the notion that the spatial distribution of matter in the universe is homogeneous and isotropic when viewed on a large enough scale.

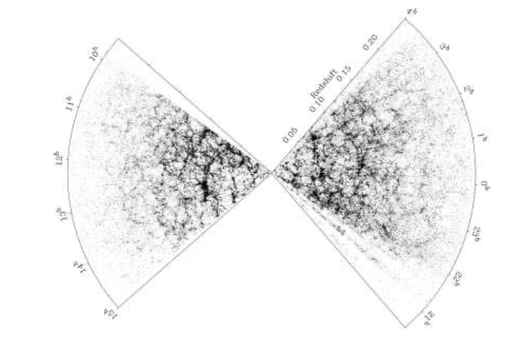


Figure 1.1: This image comes from the 2dF Galaxy Redshift Survey, which was designed to measure redshifts for approximately 250,000 galaxies, with each dot representing one of these galaxies. This survey used the 2dF multi fibre spectrograph on the Anglo-Australian Telescope, which is capable of observing 400 objects simultaneously over a 2 degree diameter field. This image is a slice through the Universe with our location being at the center of the figure, showing homogeneity among the universe as both the left and right sides of the image look nearly identical. Image adapted from Ref. [1]

In this context, "large scale" means for roughly 100 megaparsecs, where a megaparsec is about 3 million light years. Our universe dates back over 13 billion years and is said to be the result of a cosmic explosion. This explosion is known by many as the Big Bang. The Big Bang occurred everywhere in an infinite Universe and was a rapid uniform stretching of space. It is immediately following the Big Bang that small variations of density arose. Directly after the Big Bang, and for nearly 400 million years afterwards, all matter in the universe existed in an hazy state lacking any definable structure [2]. The correlations between these observations and the predictions of the Big Bang model are striking pieces of evidence in support for the theory.

The light from the Big Bang, has been going through the universe from that point forward, permitting us to recognize this "afterglow" on Earth. Our overall picture of the evolution of the Universe is that the Universe originated with small amplitude perturbations that eventually grew over time. They eventually formed galaxies and larger structures that we see around us today.

In the 1920s, Edwin Hubble attempted to measure the Doppler shift in the light from distant galaxies [3]. Doppler shift is the change in frequency of a wave in relation to an observer who is moving relative to the wave source, mostly dealt with in sound waves. Hubble observed that the light from galaxies in all directions was redshifted. Fig. 1.2 shows his findings in which you see a linear relationship between galaxy distance and velocity. He quantified this behavior as what we now call Hubble's Law, which will be defined later in Section 2.2. It was soon realized that Hubble's redshift was not caused by the Doppler effect but was instead caused

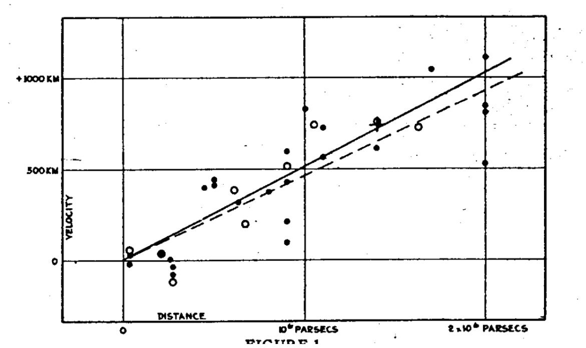


Figure 1.2: This plot is from Hubble's ground breaking paper "A relation between distance and radial velocity among extra-galactic nebulae". It shows the velocity-distance relation. The black discs and full line represent the solution for solar motion using the nebulae individually; the circles and broken line represent the solution combining the nebulae into groups; the cross represents the mean velocity corresponding to the mean distance of 22 nebulae whose distances could not be estimated individually. Image adapted from Ref. [3]

by the uniform stretching of space; Hubble had discovered the expansion of the Universe. As light travels through space that is stretching, its wavelength is also stretched. To distinguish this effect from the Doppler effect caused by motion it is called cosmological redshift.

To better explain the expansion of the Universe, our assumption is that all motion is due to gravity pulling matter from low density regions to higher density regions. The velocity due to gravity, which is called peculiar velocity, is important due to the fact that it informs us about the continuing growth of structure. Thus, with peculiar velocities we are able to see how large scale structures are growing.

In order to calculate the peculiar velocity, v, we need to know a galaxy's redshift, cz and its distance away from us, r. The distance is measured through the distance modulus, μ. As a directly measured quantity the distance modulus has Gaussian distributed errors; however, since the distance is a nonlinear function of the distance modulus, the errors in distance will not be Gaussian.

The problem we face is that we would get a biased result, meaning that the average velocity would not be correct. These non-Gaussian errors lead to biased distance measurements, which in turn lead to biased velocities. This will be talked more about in Section 2.3, but since we are trying to measure correlations of velocities as a function of separation, if the error in the separation depends on velocity there might be a bias introduced into our results. This suggests the idea that we need a different way to calculate peculiar velocities and bulk flows, which is the main premise for this thesis.

In this thesis we will be using two different methods for calculating peculiar velocities, one that is a formula and another one that is based upon a probability distribution. We will be comparing the two methods to see which one can better estimate velocity. Features of this would be a result that is unbiased and having the smallest possible uncertainty. We will also be comparing the two methods to one another to see which works better in the analysis of velocities. In order to test our hypothesis, we will use artificially generated catalogs from the two methods. The datasets we use are artificial datasets from the Outer Rim simulations. Using the artificial data allows for us to test these methods and see how well they are compared to the estimations, and finally show us if one method is better than the other for providing an accurate measurement of velocity and bulk flow.

This paper will feature the following sections: Background, Methods, Results, and Discussion. The Background section will introduce the concepts of the methods in which I worked with for this thesis. The Methods section will outline the methods used, the data sets that we used, how we used them, and why we used them in the manner that we did. The Results section covers our findings and the Discussion section goes over those results, what we found, what we learned, and directions for future work.

2 Background

2.1 History of the Universe

The Universe is said to date back around 13 billion years. Since it is impossible to see back to the birth of our Universe, we use theoretical models to study the early stages. The distribution of matter in the Universe was not perfectly uniform across, meaning that there were certain regions of the Universe that started out slightly denser than others. Over time, the matter in lower density regions was pulled towards the higher density regions due to gravity even as the average density was decreasing due to expansion.

Our Universe began with an explosion of space itself, the Big Bang. Starting from extremely high density and temperature, space expanded, the universe cooled, and the simplest elements formed. Gravity gradually drew matter together to form the first stars and the first galaxies. Galaxies eventually collected into groups, clusters, and superclusters. The cosmological structure we see today is due to galaxies moving from low density to high density regions. Regions of enhanced density originally expanded along with the rest of the Universe and due to the pull of gravity, these regions' expansion was gradually slowed down [2]. By studying peculiar velocities, we can reconstruct how mass was distributed at earlier points in time, and thus how mass fluctuations grow.

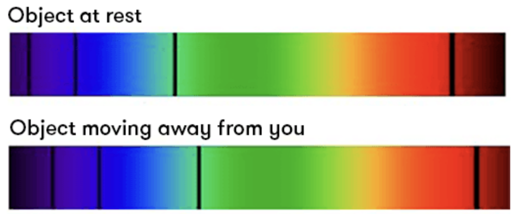


Figure 2.1: This image comes from the Australian Academy of Science and demonstrates how redhsift works. Top: the light spectrum of an object at rest. Bottom: the light spectrum of that object moving away from you. Notice how the lines shift towards the red end of the spectrum. This image is a perfect representation of redshift as the light you see from an object at rest and a moving object will not be the same. As the object moves away from you, the spectrum gets closer to the red [4]

.

2.2 Redshift

If you have ever been to a race track or watched a video of cars racing, you have probably noticed that there is a dramatic drop in pitch that happens the moment a car screams past you. You have experienced the Doppler effect. The Doppler effect refers to the difference between the frequency a detector measures and the frequency a wave source actually produces. The difference is due to the relative movement of source and detector. The phenomenon was first noticed with sound waves, but applies to all kinds of waves (including light waves). The Doppler shift is not easily detected in a continuous spectrum and cannot be measured accurately in such a spectrum. The wavelengths of the absorption lines can be measured accurately; however, most astronomical objects emit and absorb radiation at specific wavelengths called lines that occur. The wavelengths of these lines can be determined precisely making accurate measurements of the Doppler shift possible. Absorption lines are usually seen as dark lines, or lines of reduced intensity, on a continuous spectrum. This is shown in Fig. 2.1.

One of the most important concepts in measuring velocities is the Doppler effect for light. The idea of this is that if a galaxy is moving away from us, the wavelength λ_r of the light we receive from it will be longer than the emitted wavelength λ_e. If a galaxy is moving towards us, the wavelength of the light emitted will be shorter than than the emitted wavelength. We can quantify the change in the wavelength through the redshift z, defined by

$$z = \frac{\lambda_r - \lambda_e}{\lambda_e}. \tag{2.1}$$

The change of wavelength due to the Doppler Shift is expressed as

$$\lambda_r = \lambda_e \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}}, \tag{2.2}$$

where v is the observed object's velocity, and c is the speed of light [4]. Using the binomial approximation, $(1+x)^n \approxeq 1+nx$ for $x = \frac{v}{c} << 1$, which is true for a typical galaxy since most objects do not travel close to the speed of light, it can be shown that

$$\frac{\lambda_r}{\lambda_e} = \sqrt{\frac{1+\frac{v}{c}}{1-\frac{v}{c}}} \approx (1+\frac{v}{2c})(1+\frac{v}{2c}) \approx 1+\frac{v}{2c}+\frac{v}{2c} = 1+\frac{v}{c}, \tag{2.3}$$

where v represents the galaxy velocity with a positive value referring to the galaxy moving away and negative value moving towards us.

Now, if we substitute equation (2.1) and equation (2.3), we get

$$z = \frac{\lambda_r - \lambda_e}{\lambda_e} = \frac{\lambda_r}{\lambda_e} - 1 \approx 1+\frac{v}{c} - 1 = \frac{v}{c}. \tag{2.4}$$

Equation (2.4) is usually denoted as $cz = v$.

Objects also have a cosmological redshift due to the expansion of the Universe. Cosmological redshift is caused by light traveling through an expanding universe. The cosmological redshift is due to expansion rather than motion and is given by

$$cz = H_0 r, \tag{2.5}$$

where r is distance and H_0 is the Hubble constant, which describes the rate at which the universe expands.

To get the peculiar velocity we need to subtract off the cosmological redshift from the total redshift. It should be noted that there is an implicit challenge in this equation due to the fact that the total redshift is dominated by the cosmological redshift for all but the most nearby galaxies, making it difficult to extract peculiar velocity accurately. Also, distances have large uncertainties, which lead to even larger uncertainties for velocities, making them difficult to analyze. Since typical peculiar velocities are thought to be around 500 km/sec, we see that the for $H_o \approx 70\mathrm{km} \cdot \mathrm{s}^{-1} \cdot \mathrm{Mpc}^{-1}$ the uncertainties in peculiar velocities become of order their magnitudes for objects at distances $r \geq 35$ Mpc. This 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 is necessary to have a large sample in order to extract meaningful information.

2.3 Standard Candles

Apparent magnitude is the measure of the brightness of a celestial body in the sky [5]. The brighter the object, the lower the number assigned as a magnitude, and a higher number refers to a dimmer object. Apparent magnitude is a logarithmic measure of brightness that roughly matches the way our eyes perceive brightness. Apparent magnitude, denoted as m, is given by

$$m = 2.5 \log_{10} \frac{I_A}{I_B} \tag{2.6}$$

where I_A is the measured flux on Earth of the object and I_B is a reference flux. Absolute magnitude, which is denoted by M, is equal to to the apparent magnitude that the object would have if it were viewed from 10 parsecs away. Absolute magnitude is thus a measure of the luminosity of an object. Magnitudes are useful for us because if we know both the apparent and absolute magnitudes, we can calculate the distance.

Measuring the distance of galaxies is something that is not straightforward and cannot be calculated directly. To calculate the distance r, we must use what is known as the galaxy's distance modulus, μ, which is logarithmic in nature. A standard candle is an astronomical object that has a known absolute magnitude, denoted with M [5]. This is an important relation since the distance that we get from using it will be redshift-independent, meaning all variables are independent of one another. Standard candles are extremely important to astronomers since by measuring the apparent magnitude of the object we can determine its distance modulus using the formula:

$$\mu = m - M = 5 \log_{10} r - 25, \tag{2.7}$$

where m is the apparent magnitude of the object, M is the absolute magnitude of the object, and r is the distance to the object from us.

We can then calculate the distance from Equation 2.7 by

$$r = 10^{\frac{\mu}{5} - 5}, \tag{2.8}$$

with r measured in megaparsecs.

The measurements of the distance modulus have Gaussian distributed uncertainties. The fact that the uncertainty in μ is Gaussian distributed means that errors in r will not be Gaussian distributed.

The goal is then to estimate the distance, represented by r, and the peculiar velocity, represented by v, of a galaxy's given measurements of both its redshift and distance modulus. We measure the distance modulus because it is directly related to how we measure distances. It is just a fact that it has Gaussian errors

[6]. However, when we convert distance modulus to distance we find that the errors in distance are not Gaussian distributed, and this is what leads to bias. This then leads to a bias in calculating our peculiar velocity using the traditional approach. With this in mind, a reason for testing the different methods of calculating peculiar velocity is to account for this bias and try and reduce it as much as we can. This will be discussed more in the following sections.

2.4 Peculiar Velocity

Peculiar velocities are caused by the effects of gravity, which means that measurements of these velocities can be used as a probe of the large scale matter distribution. These velocities can give us insight on early mass distributions. Having an idea of early mass distributions then leads us to have a better picture of what the early Universe could have been like with density fields [7]. It would be like looking at fossils of dinosaurs on Earth and what that tells us about that time in history.

In order to measure the peculiar velocity of a galaxy, we need both the galaxy's redshift cz and distance, r. We need the distance so we can subtract off the cosmological redshift and find the part of the redshift that is due to motion. The movement caused only by the effects of gravity on a galaxy are different for each galaxy we observe, and so their velocities are unique/peculiar. The velocity can be calculated with the formula

$$v = cz - H_0 r, \tag{2.9}$$

where cz is the galaxy's redshift, r is its distance, and H_0 is Hubble's constant. What this formula is doing is taking the overall redshift of cz and subtracting or taking out the effect that Hubble's expansion $H_0 r$ has.

Having peculiar velocities is important for a few reasons. The first reason is that it will allow for us to have a data set were we actually know the peculiar velocities roughly. With a catalog of cz and a distance modulus μ catalog, we can calculate the corresponding velocities and examine the concerned galaxies' behavior on small scales. Having individual peculiar velocity estimates will lead to larger uncertainties, but with a larger pool of data, and averaging over this pool, it allows for us to have smaller uncertainties and more reliable data to work with.

The second application of peculiar velocities is that they allow for us to calculate larger scales flows, also known as bulk flows. Bulk flow is the average velocity for a group of galaxies. These flows are evidence that the universe is still expanding and that large scale structure is still forming [8]. A central idea of bulk flow is to average over volumes; while each velocity has a very large uncertainty, the average velocity should have much smaller uncertainty. The bulk flow can then be compared to predictions of the Big Bang Model. Homogeneity and isotropy of the Universe tells us that the bulk flow should go to zero if we consider a large enough volume; thus bulk flows can allow us to test if the Universe satisfies the Cosmological Principle [9].

3 Methods

This section will include a detailed explanation of both the formula method and probability method that we used in this work. The mathematics needed to understand each method will be addressed in this section.

3.1 Probability Method

The first method in which we will try to accurately calculate peculiar velocity is based upon a probability distribution. It is important to note that this probability distribution does not account for individual measurement of velocities, but rather averages over a large number of measurements.

The distance modulus μ corresponding to a given distance r (in Mpc) is given by

$$\mu = 5\log_{10}(r) + 25 + \delta, \tag{3.1}$$

where δ is measurement noise drawn from a given Gaussian distribution centered around 0. The measurement uncertainty in μ is represented by σ_μ, which is given for each object. Measurements of the distance modulus have absolute uncertainties mostly around 0.5, meaning a the typical value is $\sigma_\mu = 0.5$. Taking the inverse of equation (3.1) we have,

$$r = 10^{\frac{\mu-\delta}{5}-5}. \tag{3.2}$$

Some clusters of galaxies have multiple measurements and have uncertainties smaller than 0.1. The fact that the uncertainty in μ is Gaussian distributed means that errors in r will not be Gaussian distributed.

The goal is to calculate the bias in distance measurements. Since the measured μ is related to the true μ_o through $\mu = \mu_o + \delta$, we can calculate the measured distance as

$$r = 10^{\frac{\mu}{5}-5} = 10^{\frac{\mu_o}{5}-5}10^{\frac{\delta}{5}} = re^{\frac{\delta \log 10}{5}}, \tag{3.3}$$

where we have used the fact that $10 = e^{\log(10)}$.

Given that the probability distribution for δ is

$$P(\delta) = \frac{1}{\sqrt{2\pi\sigma_\mu}} e^- \frac{\delta^2}{2\sigma_u^2} \tag{3.4}$$

we can write the average of the measured values of a particular galaxy distance r over many realization as

$$\langle r \rangle = r_o \frac{1}{\sqrt{2\pi\sigma_\mu}} \int_{-\infty}^{\infty} e^- \frac{\delta^2}{2\sigma_u^2} d\delta \approx r_o \frac{1}{\sqrt{2\pi\sigma_\mu}} \int_{-\infty}^{\infty} e^{\frac{\delta \log 10/5 - \delta^2}{2\sigma_u^2}} d\delta \tag{3.5}$$

where r_o is the actual distance. We can then do the Gaussian integral by the method of completing the square, which results in

$$\langle r \rangle = r_o e^{\frac{\log 10\, \sigma_\mu^2}{50}}. \tag{3.6}$$

Therefore we can correct for the bias by dividing our distance estimates calculated from the distance modulus by a factor, so that distance estimates given by

$$r_e = \frac{10^{\frac{\mu}{5}-5}}{e^{\frac{\log 10\, \sigma_\mu^2}{50}}}, \tag{3.7}$$

are unbiased in that $\langle r_e \rangle = r_o$.

We can then calculate peculiar velocities using

$$v = cz - H_o r_e \tag{3.8}$$

with the corrected distances and the resulting velocity estimates will also be unbiased.

3.2 New Estimator for Peculiar Velocity

The next method in which we will discuss is a formula method for calculating the peculiar velocities. This formula estimates the peculiar velocity of a galaxy or group of galaxies from redshift and distance modulus. It is called the new estimator for peculiar velocity developed by Richard Watkins and Hume A. Feldman from "An Unbiased Estimator of Peculiar Velocity with Gaussian Distributed Errors for Precision Cosmology" [6].

As stated previously, the most straightforward estimator is given by

$$v = cz - H_o r, \tag{3.9}$$

where v it the velocity, cz refers to the redshift, H_o is Hubble's constant, and r the distance. This is typically the estimator used in peculiar velocity analyses.

Equation (3.9) shows the most straight forward estimator used in peculiar velocity analyses. This formula uses the log distance, or distance modulus, since it has Gaussian distributed errors. Taking the exponential will then lead to a skew in our error distributions, resulting in distance errors that are not Gaussian distributed. Another issue with Eq. (3.9) is that the average of a group of velocity estimates with different errors is not the true value, meaning that $\langle v_e \rangle \neq v$. This is the result of the skewness of the distribution of distance errors, which gives $\langle r_e \rangle \neq r$. Thus, with these issues in mind, Watkins and Feldman propose the new estimator as,

$$v_e = cz\ln(cz) - cz((\ln(10)(\frac{\mu - 25}{5})) + \ln(_H 0)) \tag{3.10}$$

The goal of Eq. 3.10 is to obtain an estimate v_e of the peculiar velocity of a galaxy or group from the galaxy's redshift, denoted as cz, and an estimate of it's distance r_e. Using the log of the distance estimates is an important feature as it has Gaussian distributed errors. An important note for this formula is that it has already been proven that it will only work based upon a limit, meaning that the limit fails for nearby objects. Using $\langle \log(r_e) \rangle = \langle r \rangle$, they show that the estimator is unbaised as long as the true $v << cz$. Watkins and Feldman break this down through

$$v_e = cz\ln(cz) - cz((\ln(10)(\frac{\mu - 25}{5})) + \ln(H_0)), \tag{3.11}$$

where we first rewrite the second term in terms of r instead of μ and pull out the common factor of cz.

$$v_e = cz(\ln(cz) - \ln(H_0 r)) \tag{3.12}$$

Now we write

$$\ln(H_0 r) = \ln(cz - v) = \ln(cz) + \ln(1 - \frac{v}{cz}). \tag{3.13}$$

If $v \ll cz$, we can use the fact that $\ln(1 + x) \approx x$ for $x \ll 1$ to write

$$\ln(1 - \frac{v}{cz}) \approx -\left(\frac{v}{cz}\right). \tag{3.14}$$

Plugging this back in we get

$$v_e = cz(\ln(cz) - \ln(H_0 r)) \approx cz\left(\ln(cz) - \ln(cz) + \left(\frac{v}{cz}\right)\right) \approx v, \tag{3.15}$$

Assuming that the uncertainties in the redshift, cz are negligible.

3.3 Bulk Flow

Cosmological bulk flows are the average of peculiar velocities over spherical regions around us. Bulk flows are usually considered for sufficiently large spheres where linear expressions for the velocity and density power spectra are valid. Since the Universe is homogeneous and isotropic, the bulk flow should go to zero as the radius becomes large. The bulk flow can then be compared to predictions of the Big Bang model. The concept of bulk flow is fairly easy to understand, but the measurements of bulk flow can be difficult to interpret. The reason for this is that when dealing with peculiar velocity measurements, we have uncertainties that grow rapidly with distance. There is also the fact that with most catalogs, nearby galaxies are overrepresented as compared to further ones. This can lead to the bulk flow measurements being reflective of scales somewhat smaller than that of the survey.

For this project specifically, we measured bulk flow using both methods and comparing and analyzing the two. In order to do this, we had to calculate the bulk flows of 300 catalogs from the Outer Rim Simulation. This could usually be done by taking the average velocity over many individual measurements to determine the general flow of galaxies within a specific region. The issue with this is that this will not result in uniform sampling. For example, galaxies that are closer to a given catalog center have more weight in determining the overall bulk flow of a catalog than far away galaxies, which biases our large scale results.

A way to envision bulk flow is by thinking of the average of a full, three-dimensional peculiar velocity, represented by v_i over a spherical volume V with a radius R:

$$U_i = \frac{1}{V}\int_V v_i d^3 r, \tag{3.16}$$

where $i = x, y, z$ are the Cartesian components of the velocity field [9]. The weights are adjusted so that each part of the volume gives the same contribution. In other words, a larger weight would mean that there is objects in regions where there are fewer object,s and a smaller weight would mean that there is objects in regions with a lot of objects.

For the nonuniform sampling we can weight each individual velocity. t may not always result in uniform sampling, we can add weights to specific galaxy velocities. By doing this, we can simulate how real data is used to perform bulk flow analysis. For the x-th directional-component of a catalog's bulk flow, U_i can be found by using

$$U_x = \sum_i^N w_i v_{ix}, \tag{3.17}$$

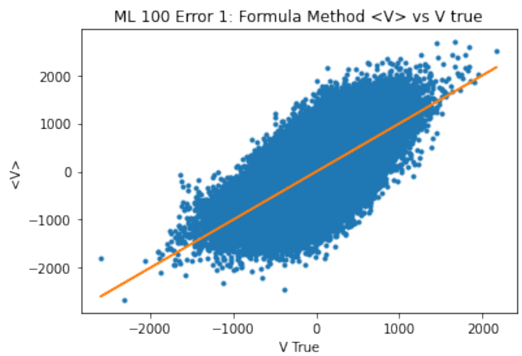
where v_{ix} is the x-th directional component of the i-th velocity estimate in the catalog we are using and w_i is the corresponding weight, which is calculated through a series of codes [9]. The weights account for how much radial velocity in a given direction will contribute to a certain bulk flow component. A way to think about this is that objects in the x-direction will mostly contribute to the x-component bulk flow and will have small weights in the sums to calculate the y and z components of the bulk flow [9]. However, these bulk flows are not perfect and have uncertainty, meaning velocities with larger uncertainties should generally get less weight in the sums. We can then use Equation (3.17) to calculate the three components x, y, and z of the bulk flow for a given catalog.

4 Results

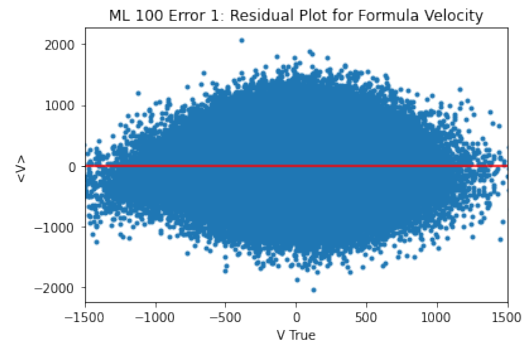
4.1 Estimating Peculiar Velocities Results

In this section, we will go over the results we obtained from the two different methods for calculating peculiar velocities. Along with that we will also go over the results we obtained from the bulk flow analysis also using the two methods.

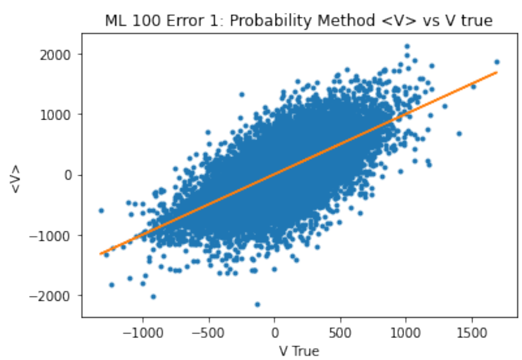
When looking at Fig. 4.1a and Fig. 4.1c we can see that both methods are estimating peculiar velocities with good accuracy. The orange line on each graph shows what the true velocity should look like. If the methods were to perfectly be estimate velocity, we would expect for all of the blue points to be on the orange line. Even though this is not the case, the methods are estimating fairly well, given that the distribution of points above and below the orange line are in an unbiased way. The skewness of the graphs are reasonable, given that we are working with a lot of data. As the uncertainty increases, we see that the overall skewness of the graphs increase as well. This means that as we increase the uncertainty value σ_μ, this will lead to a less accurate estimation of velocity for both methods. When looking at the residuals for both methods, we see that the velocities are unbiased and have the uncertainties that we would expect due to the fact that the residual plots averages are very close to 0. These residuals tell us how far off from the expected velocities our estimates are. We can then assume that both methods are unbiased and, when calculating large-scale galaxy movement, will give accurate answers.



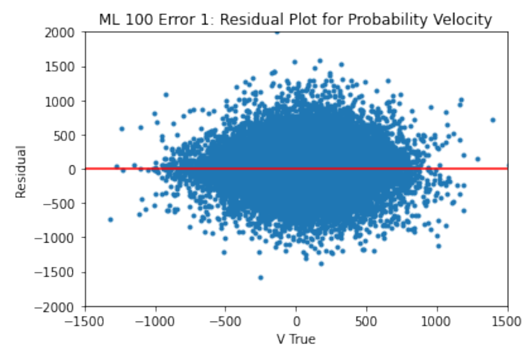
(a) Formula Velocity vs True Velocity for uncertainty of μ at 0.1



(b) Residual Plot of Velocity Calculated from Formula Method

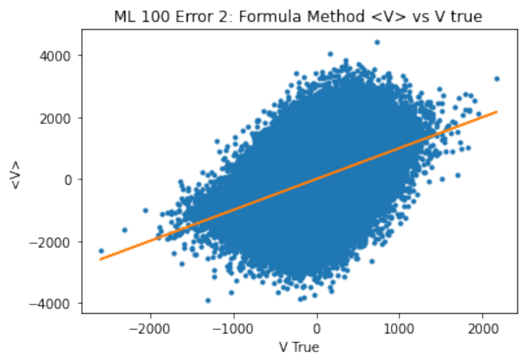


(c) Probability Velocity vs True Velocity for uncertainty of μ at 0.1

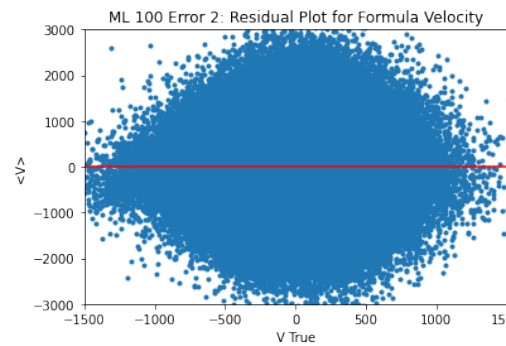


(d) Residual Plot of Velocity Calculated from Probability Method

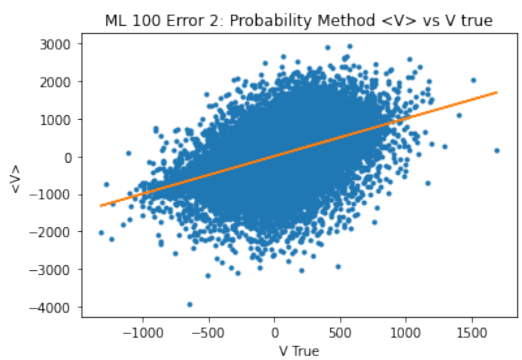
Figure 4.1: Plots (a) and (c) show the predicted velocity vs the true velocity. The true velocity comes from simulated data catalogs. The predicted velocities are estimated from both the formula method and probability methods. Plots (b) and (d) show the residuals for each of these methods.



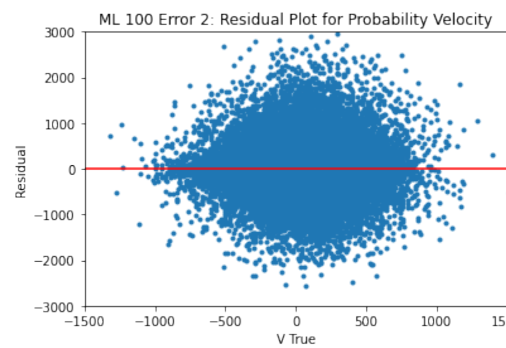
(a) Formula Velocity vs True Velocity for uncertainty of μ at 0.2



(b) Residual Plot of Velocity Calculated from Formula Method

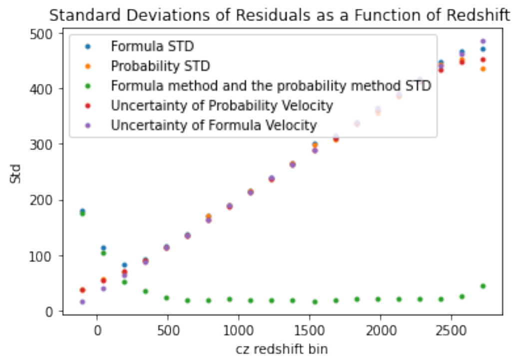


(c) Probability Velocity vs True Velocity for uncertainty of μ at 0.2

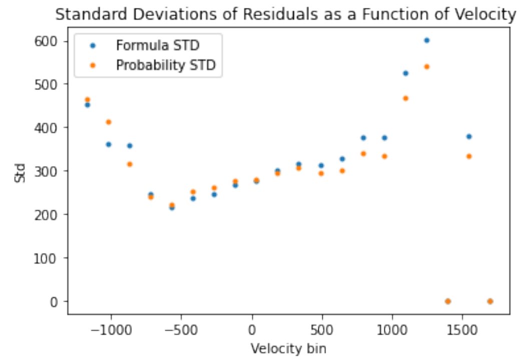


(d) Residual Plot of Velocity Calculated from Probability Method

Figure 4.2: Plots (a) and (c) show the predicted velocity vs the true velocity. The true velocity comes from simulated data catalogs. The predicted velocities are estimated from both the formula method and probability methods. Plots (b) and (d) show the residuals for each of these methods.



(a) Standard Deviation of Residuals as a Function of Redshift at $\mu = 0.1$

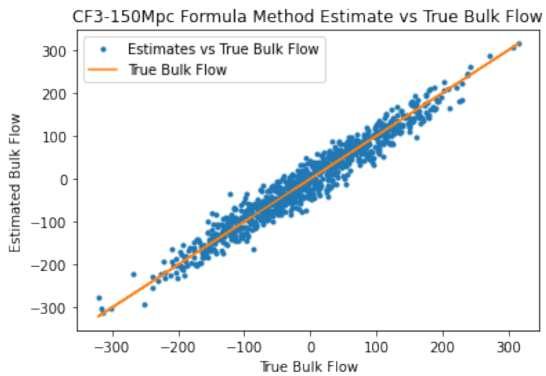


(b) Standard Deviation of Residuals as a Function of Velocity at $\mu = 0.1$

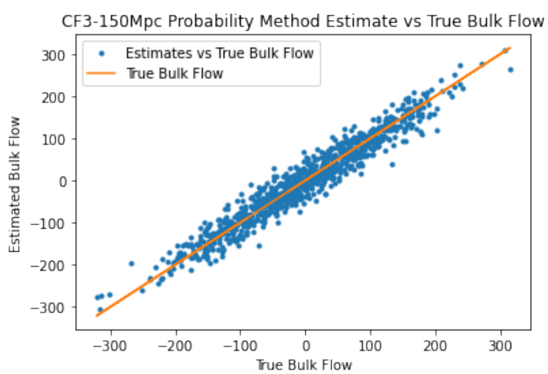
Figure 4.3: These graphs show the standard deviations of the residuals as a function of redshift for both the formula and probability distribution methods. The redshift was binned in order to show a more clear graph and depiction of what exactly is happening. We can see that the point, $cz \approx 270$, in which formula method seems to have a lower standard deviation of residuals than the probability method.

4.2 Standard Deviations of Residuals as a Function of Redshift and Velocity

As we begun looking at the analysis for both methods, there was not too much difference overall between the two as the average of the residuals were near zero for both methods. So, to the determine if there were any differences, we decided to look at the standard deviation of the residuals for the estimators. Fig. 4.3a and Fig. 4.3b show the standard deviation of residuals for the true and estimated velocity values. What we got from this was a confirmation of the fact that the formula method does not work well for small redshift values. From the graph, we can see that the formula method, regardless of the uncertainty value, always under produced, or gave velocity measurements with more error at small redshift values as opposed to the probability method. Then, at a certain redshift, both methods would produce similar results. Another find, was that as we increased the uncertainty of μ, the point in which the formula method was under performing actually decreased. So for a bigger value of μ there is not a big variation between the two methods.

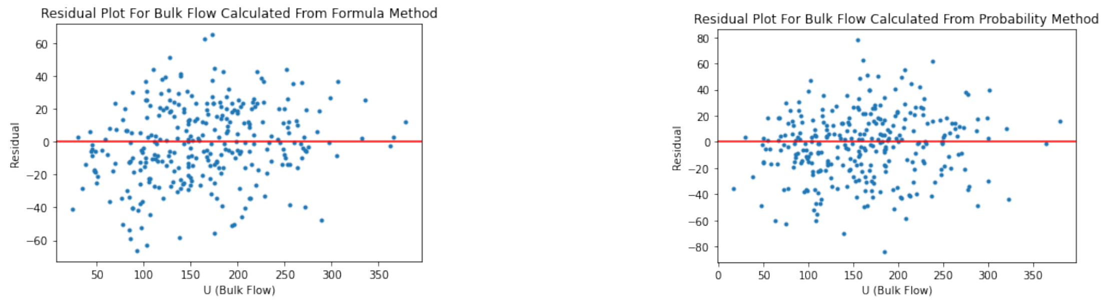


(a) Plot of the estimated bulk flow vs the true bulk flow for the formula method.



(b) Plot of the estimated bulk flow vs the true bulk flow for the probability method.

Figure 4.4: These graphs show the plot of true bulk flow vs the estimated bulk flow from the two methods, formula and probability method. The orange line represents the true bulk flow, which is predicted using simulated data. The blue dots are each estimated bulk flow. Based on side by side comparison, we can see that both methods are producing similar results as the correlation of each graph is very similar.



(a) Bulk Flow Residual Plot Estimated from Formula Method

(b) Bulk Flow Residual Plot Estimated from Probability Method

Figure 4.5: These graphs show the residual plots for the bulk flow calculated using both the formula and probability method. This is the residuals for the total bulk flow, meaning that all of the components, x, y, and z are added up. The average of the residuals for both are near zero, meaning that there is no major difference between them.

4.3 Bulk Flow Results

To further assess the two methods, a bulk flow analysis in spheres of radius 150 Mpc was conducted, which allowed for us to see in more detail how well the estimates truly work. Fig. 4.4a and Fig. 4.4b are the graphs showing the plot of the true bulk flow vs the estimated bulk flow from the methods. As observed from the graph we can see that the formula method works well for estimating the bulk flow as there is a nice correlation between the data points along the true velocity line. The same is true for the probability method, as the correlation is similar above and below the line. We can also see from Fig. 4.5a and Fig. 4.5b that both the averages of the residuals for the calculated methods are still near zero. This confirms the fact that the methods are producing unbiased results and correcting the factor from the distance modulus to distance conversion.

5 Discussion

The purpose of this project was to analyze and asses two methods for estimating peculiar velocity. The first being a formula that estimates peculiar velocity and the second method is a probability distribution that accounts for the non-Gaussian distributed error in distance estimates. From using artificial data, and plotting true velocities against estimated velocities, it is apparent that both methods seem to be estimating peculiar velocity at similar rates. Their residuals are a good indicator of this as both have residual averages that are very close to zero. This tells us that they are correcting the distance bias. It is also apparent, which was stated in their work, that the “New Estimator” does not work well for small distances, and that the probability method is a better way to estimate peculiar velocity for smaller distances. This was shown in Fig. 4.1c.

When looking at the bulk flow produced from both methods, it was apparent that they are producing very similar results for bulk flow. This is a interesting find due to the fact that we were expecting some differences in the smaller distance since the formula method did not perform as well as the probability method. But despite this, the bulk flow estimates seem to work really well for the bulk flow and produced an accurate result.

A look into the future would be to see if there can be a way to curate catalogs using both the probability and formula methods, using the probability method for small distances and then either method for larger distances. We believe that this would allow for very accurate estimates of peculiar velocities. Then with accurate weights, this could lead to improvements in measurement of bulk flows.

Bibliography

[1] M. Colless, "The 2df galaxy redshift survey," *The 2dF Galaxy Redshift Survey*, Jun 2003.

[2] B. S. Ryden, *Introduction to cosmology.* Cambridge University Press, 2017.

[3] E. Hubble, "A relation between distance and radial velocity among extra-galactic nebulae," *Proceedings of the National Academy of Sciences*, vol. 15, pp. 168–173, 1939.

[4] E. Gregersen, "Astronomical redshift," *Imaging the Universe - The University of Iowa*, Apr 2018.

[5] M. Postman and T. R. Lauer, "Brightest cluster galaxies as standard candles," *Astrophysical Journal*, vol. 440, p. 28, 1995.

[6] H. Fieldman and R. Watkins, "An unbiased estimator of peculiar velocity with Gaussian distributed errors for precision cosmology," *Monthly Notices of the Royal Astronomical Society*, vol. 450, pp. 1868–1873, 2015.

[7] A. Nusser, "On methods of estimating cosmological bulk flows," *Monthly Notices of the Royal Astronomical Society*, vol. 455, pp. 178–184, 2016.

[8] A. Faith, S. Salehi and M. Yarahmadi, "Cosmological bulk flow in the QCDM model: inconsistency with λcdm," *Monthly Notices of the Royal Astronomical Society*, vol. 504, pp. 1304–1319, 2021.

[9] S. Peery, R. Watkins, and H. A. Feldman, "Easily interpretable bulk flows: Continuing tension with the standard cosmological model," *Monthly Notices of the Royal Astronomical Society*, vol. 481, pp. 1368–1375, 2018.