

Responsibly and Usefully Cherry Picking Data for Astronomy

Alex Tinat

- Title: Optimal Moments for the Analysis of Peculiar Velocity Surveys
- Authors: Richard Watkins, Hume A. Feldman, Scott W. Chambers, Patrick Gordman and Adrian L. Melott
- First author institution: Willamette University
- Status: Accepted for Publication in the Astrophysical Journal

You might be familiar with some names for the groupings of astronomical objects: star systems, galaxies, even clusters of galaxies, but some research is concerned with even bigger things, such as superclusters and the mythical “density field” and “super-duper-clusters.” We can’t look directly at these large groups, instead we can only measure specific objects and their properties, like their velocity. In order to get a sense of what’s happening on the larger scales, we combine these numbers, and the paper today looks at how to do this in a way that preserves the important information.

Our subject of astronomy is concerned with the the power spectrum of fluctuations in the density field. That was definitely a mouthful, so let’s start simple. We generally assume that the universe is homogenous, it’s pretty similar all throughout (R. Watkins, personal communication, March 8th, 2024). There are, however, some fluctuations and we can describe those fluctuations as waves. An area of space with more galaxies (or just, mass) than normal is thus a fluctuation wave, with wavelength on the order of the size of that area of space. The power spectrum of these fluctuations is a picture showing how much of each size of wave is present in the universe (if you know what a Fourier transform is, it’s very similar in essence!). At large scales, these waves do indeed obey like waves: they don’t interact with one another, and they add linearly (they follow the superposition principle).

The method this paper talks about is determining the power spectrum by looking at velocity data and using its connection to mass density [1]. The correlation between velocity and mass comes from our good friend gravity, which shows up by saying that if there’s a high amplitude fluctuation in the density field, the velocities in that region should be correlated (R. Watkins, personal communication, March 8th, 2024). Basically, if there’s more mass than usual in a spot, gravity will be stronger in that region, so things will have some pull, and thus velocity, towards the center of mass of a fluctuation. A velocity on one side of the fluctuation will be towards the center, and is thus correlated to a velocity on the other side, which will be in the reverse direction towards the center. Using this relation, it is possible to take a set of velocities, or moments (which is an all too unusual word for “weighted sum”) made from those velocities, look at their correlations, and figure out what parameters for the power spectrum would fit this data best, a kind of approach known as a likelihood analysis.

Our treatment of density fluctuations as waves only works in the big picture. On smaller scales, things get more complicated, and the model doesn’t work so well [1]. We therefore want less small scale information and more large scale data in order to apply the aforementioned theory. And with all that background covered, we can get to the meat of this paper: a method for compressing data while also separating moments based on whether they contain more information about smaller scales or larger scales.

The math, in its full glory, would be too big a mouthful for this WUbite and so only the general process will be described. Data is combined into moments by taking a linear combination, so each term is multiplied by some constant and then added together [1]. The trick this paper works out is how to pick those constants so that the moments will have varying sensitivity to nonlinear scales, and a corresponding value λ (an eigenvalue, actually!) which quantifies that sensitivity. The moments can then be arranged by their size of λ , and those highest can be excluded from an analysis that tries to identify high order density fluctuations.

This method was then used on data from a simulated universe that should look similar to ours, to see if it works as intended [1]. One check was made on this method to make sure it did two of things we wanted: exclude information on smaller scales, and not actually throw out ALL the data, including the higher order stuff. This check was done by using a window function, which was constructed to show how sensitive a given moment is to different scales of wave information. As you can see in Figure 1, moments with small λ do only show sensitivity to higher order (low wavenumber) power, and vice versa, which is exactly the desired effect. The likelihood analysis was then carried out, to see how well we can predict the parameters of the power spectrum using this method, and how impactful the data compression and selection methods are. As you can see in Figure 2, throwing out some moments, starting with those with the highest sensitivity to small scale power, increased agreement between the predicted and true value. The process works just as hoped!

This paper was published in 2002 (before I was born!), and a lot can be said since its release. The process seems successful, and the data compression method is valid for other types of large scale probing [1]. The use of this method in the field, however, is sparse. The power spectrum parameters this paper was concerned with have been accurately calculated using distinctly different methods, unrelated to velocity survey data (R. Watkins, personal communication, March 8th, 2024). While it might seem underperforming right now, the process this paper outlines is rumored to be used in the next great astrophysics paper to come out of Willamette University, so don't give up on it just yet!

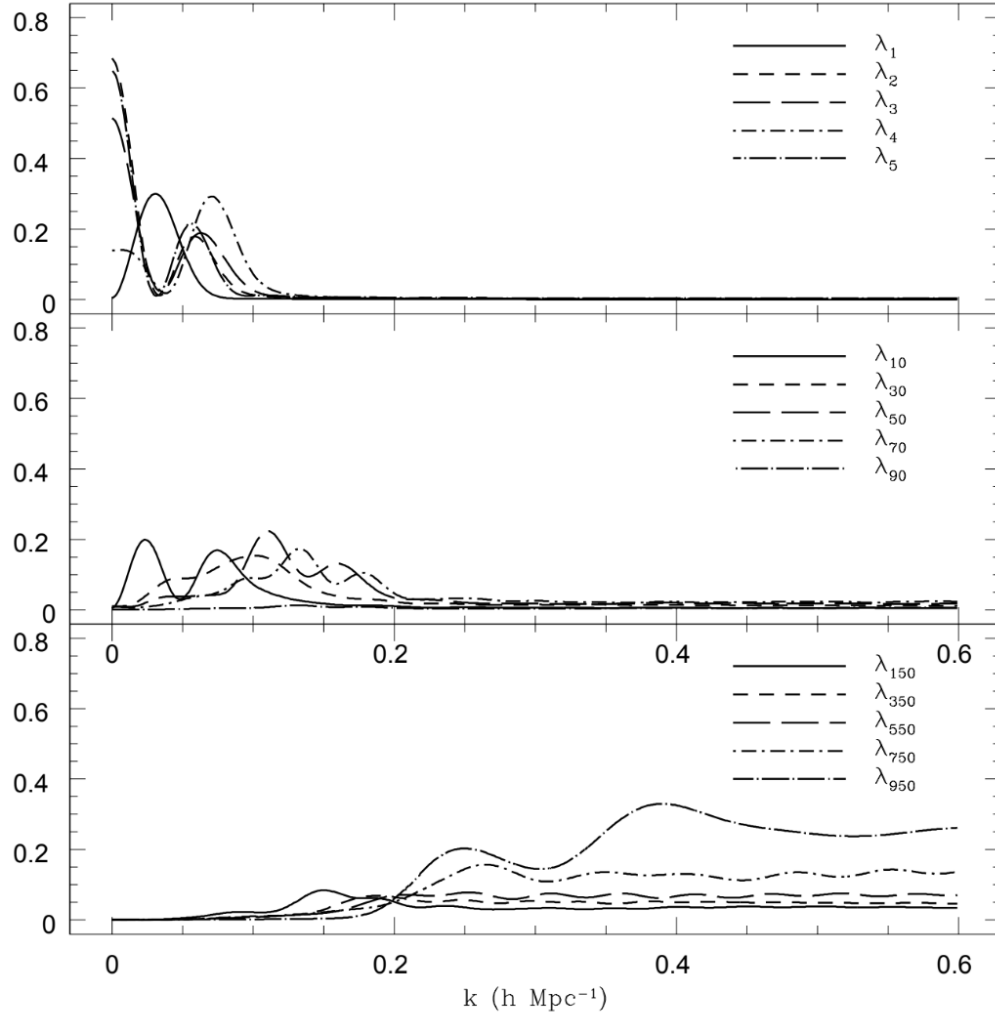


FIG. 1. Window functions of moments with varying sizes of λ . Shows how sensitive a given moment is to different scales of power[1].

[1] R. Watkins, H. A. Feldman, S. W. Chambers, P. Gorman, and A. L. Melott, Optimal moments for the analysis of peculiar velocity surveys, *The Astrophysical Journal* **564**, 534 (2002).

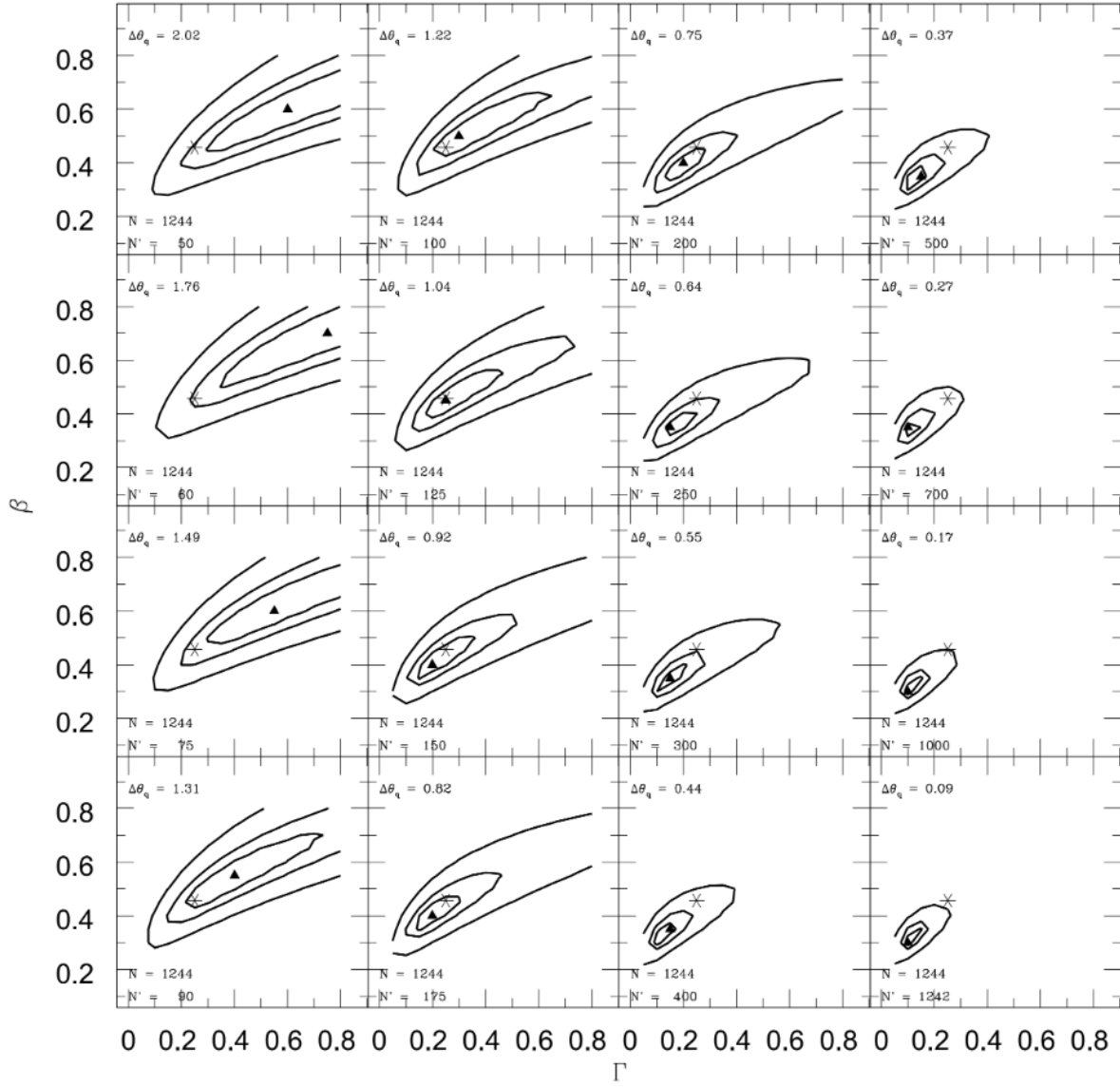


FIG. 2. Likelihood contour maps for the fit of power function function parameters. Different plots have varying numbers of kept moments. The star is at the true value from the universe simulation creation [1].