



# RESEARCH UPDATE IV

10-620 | Independent Study: Research | Andrea Klein

# PROBLEM SETUP

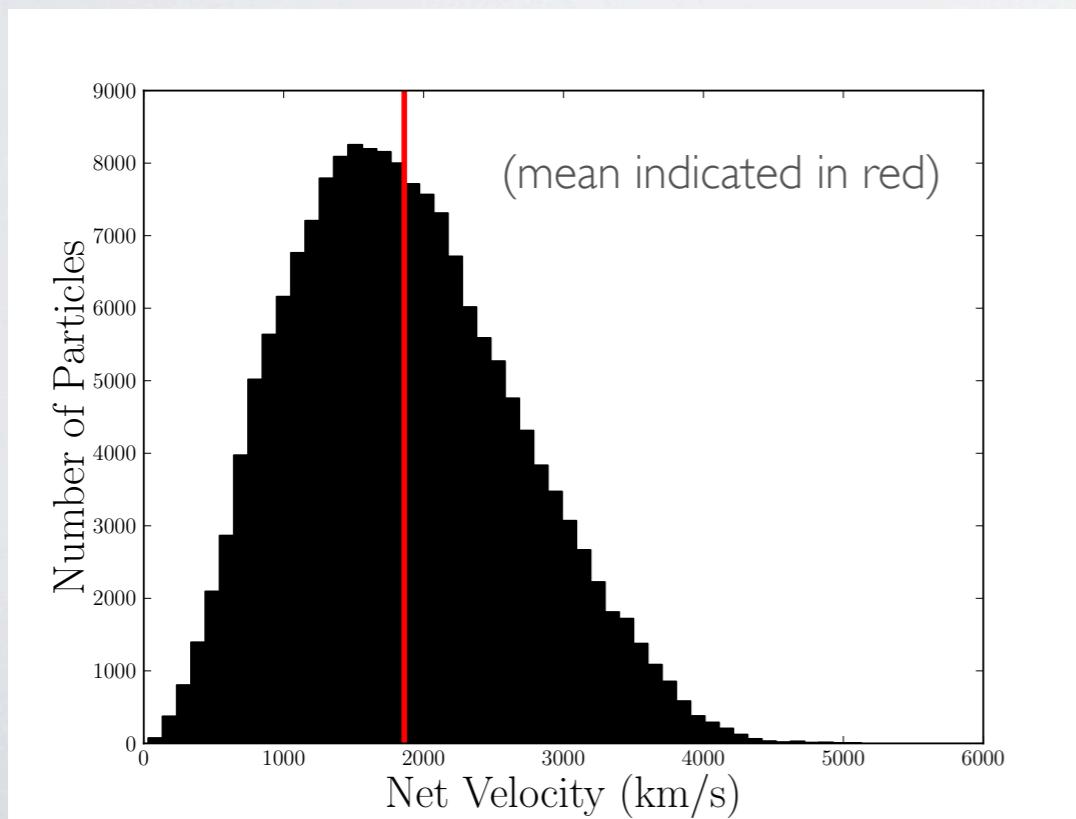
- Recall that I am attempting to learn the following two functions (separately) via SVR:
  - [ $\langle \text{Mass}, \text{R200a}, \text{Mcmax}, \text{Rcmax} \rangle$ ] -> average net velocity
  - [ $\langle \text{Mass}, \text{R200a}, \text{Mcmax}, \text{Rcmax} \rangle$ ] -> std. dev. net velocity
- I have determined that the radial positions of the particles are not informative, so I'm just using halo parameters (that characterize the halo's size and shape) as input. The net velocities, however, are taken from the particle data.
- For future work with subhalo velocities, I'd likely use  $\text{R/Rvir}$  as additional input.

# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest

# JUSTIFICATION

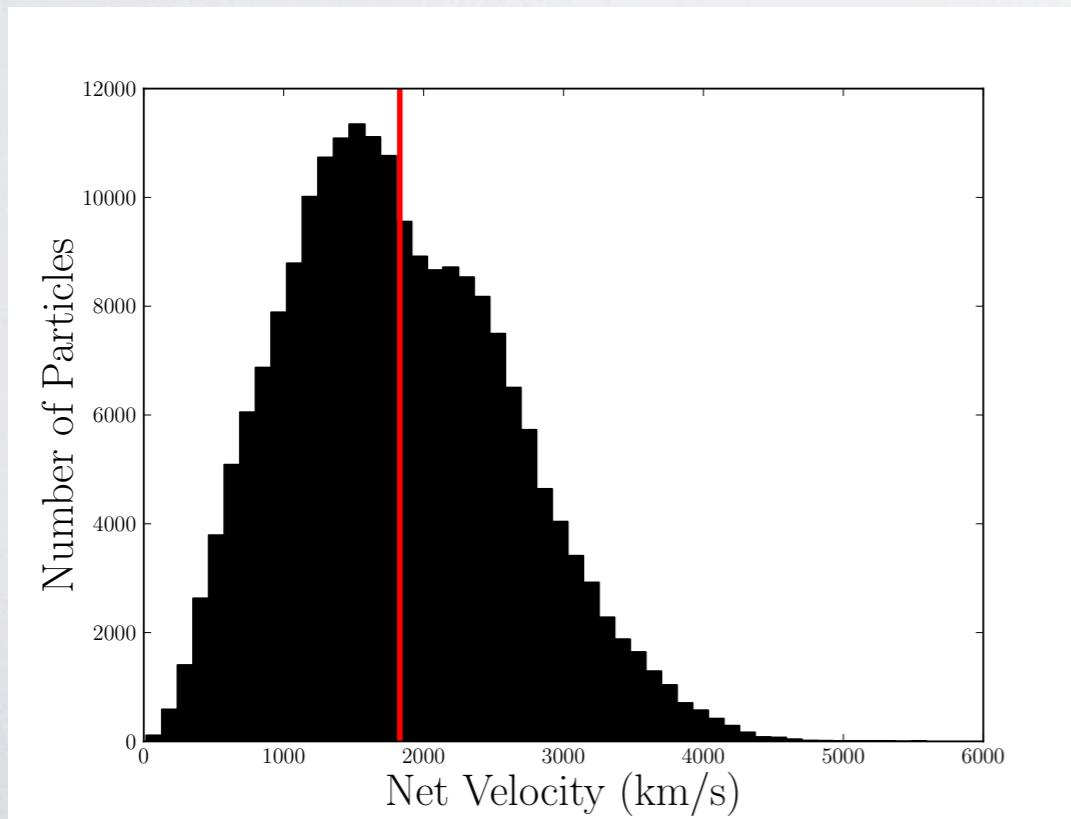
- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest



Most of the velocity distributions look about like this: roughly normal, but with significant skew.

# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest



Some have double-peaks  
due to recent mergers.

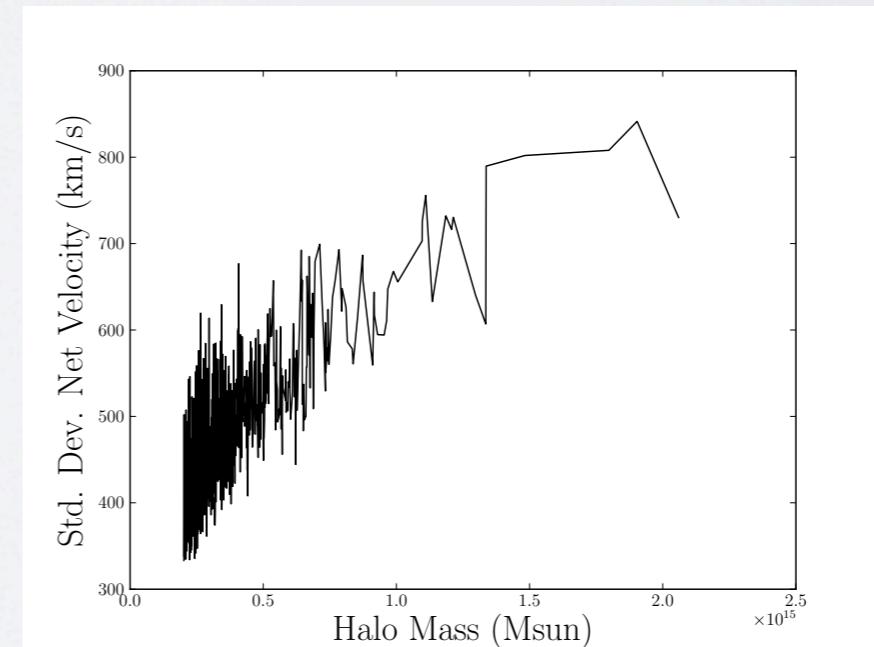
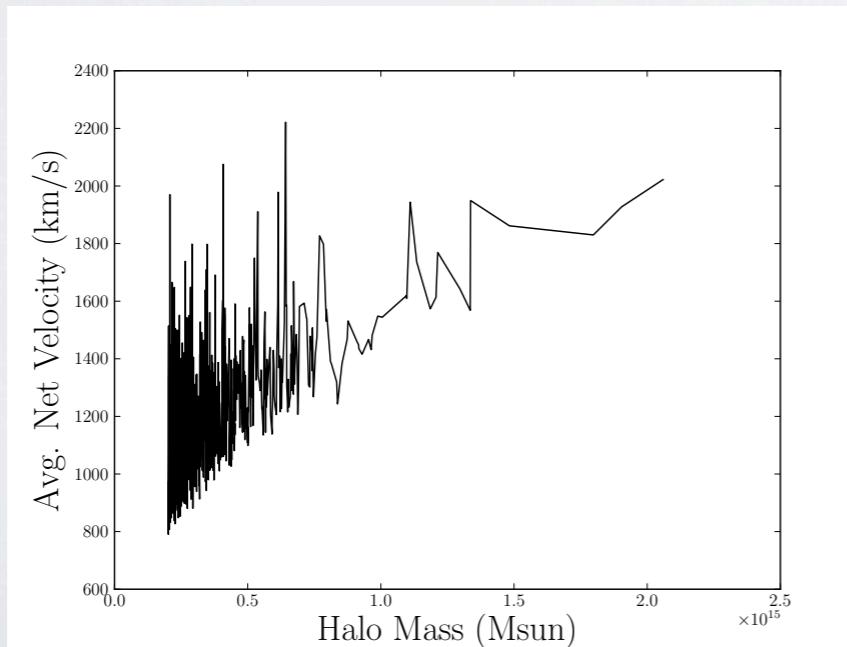
# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest

Unfortunately, a Kolmogorov-Smirnov test (on normalized Vs) yields negligible p-values. So we must take the normal distribution as only a rough parametrization of the true distribution.

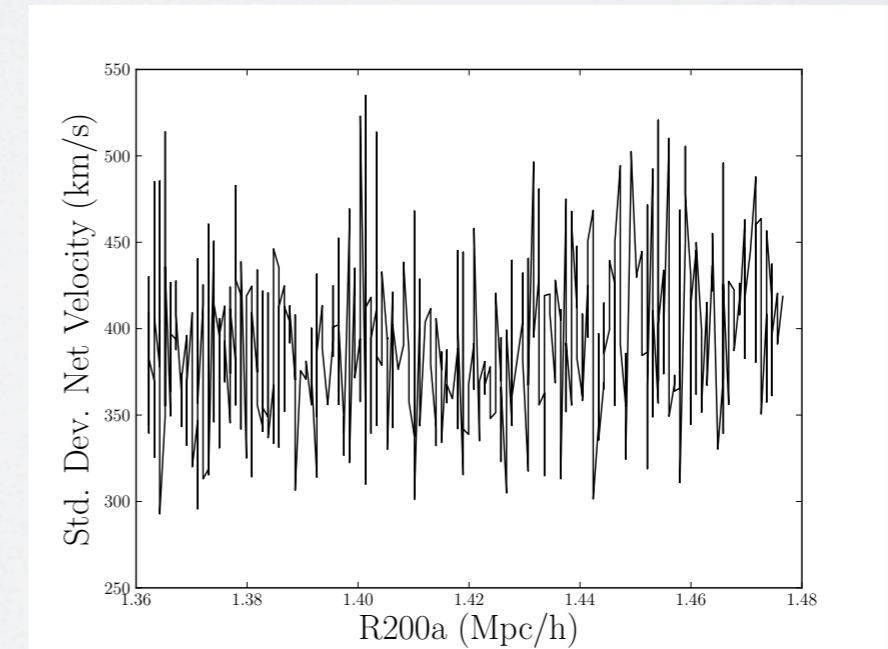
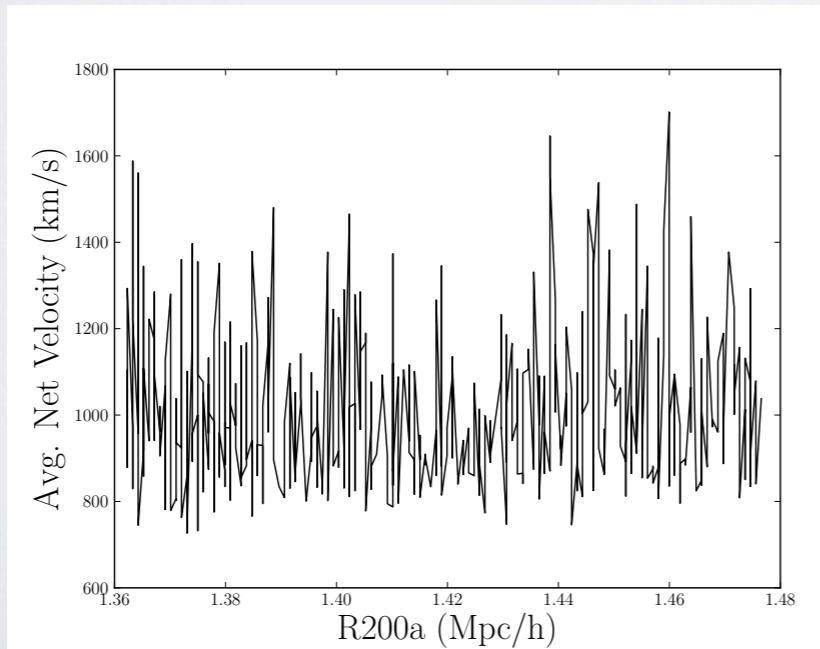
# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest



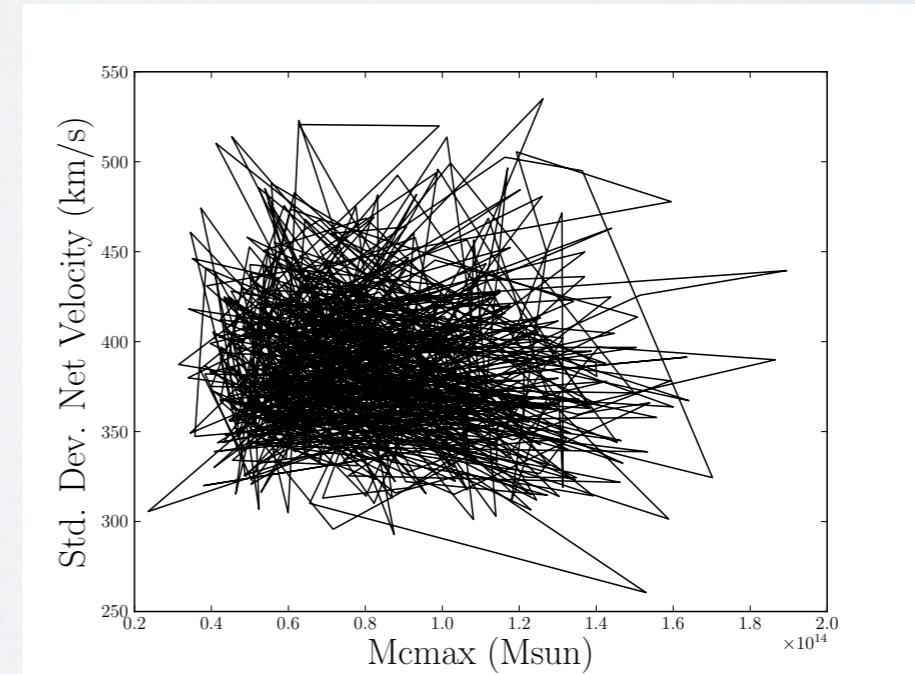
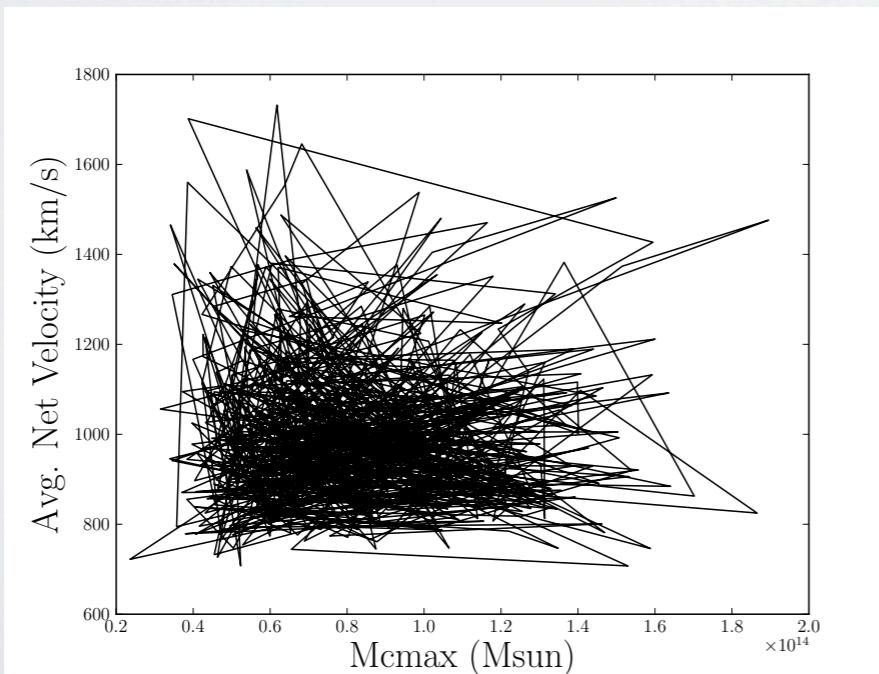
# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest



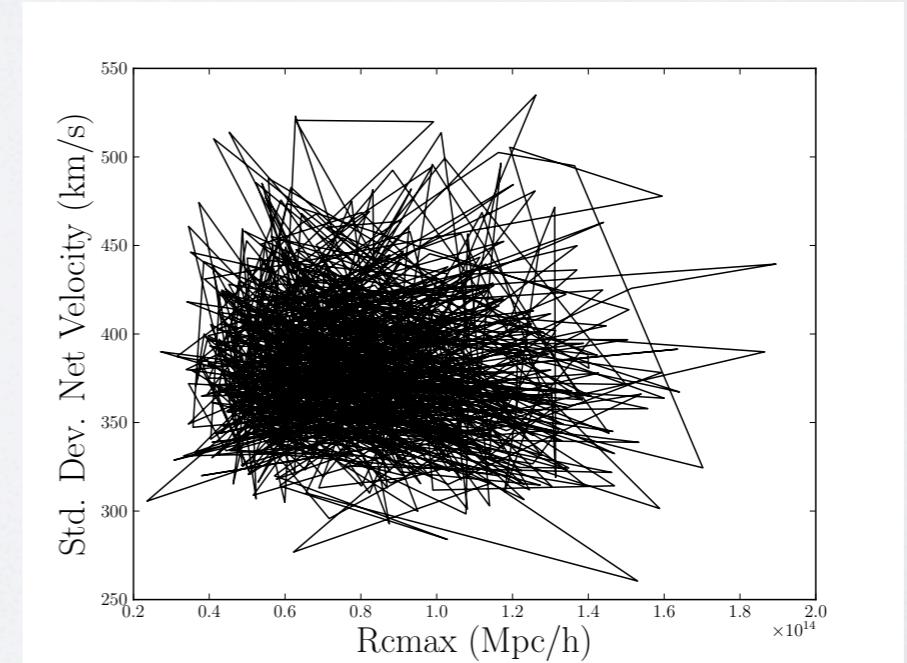
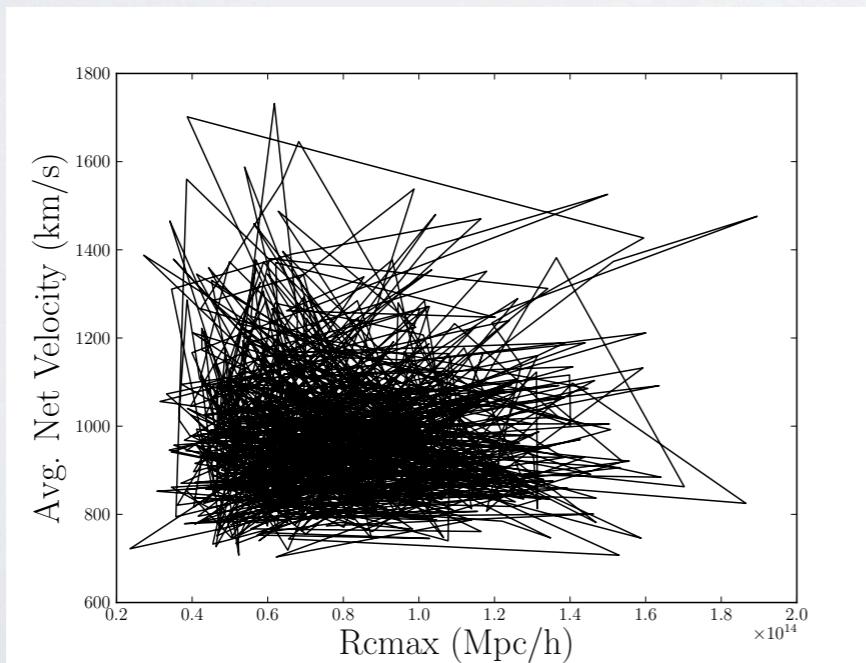
# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest



# JUSTIFICATION

- In order to justify this setup I ought to demonstrate two things:
  - (1) the velocities for a given halo are normally distributed, so the entire (net) velocity distribution can be characterized in this way
  - (2) The input parameters are good predictors of the quantities of interest

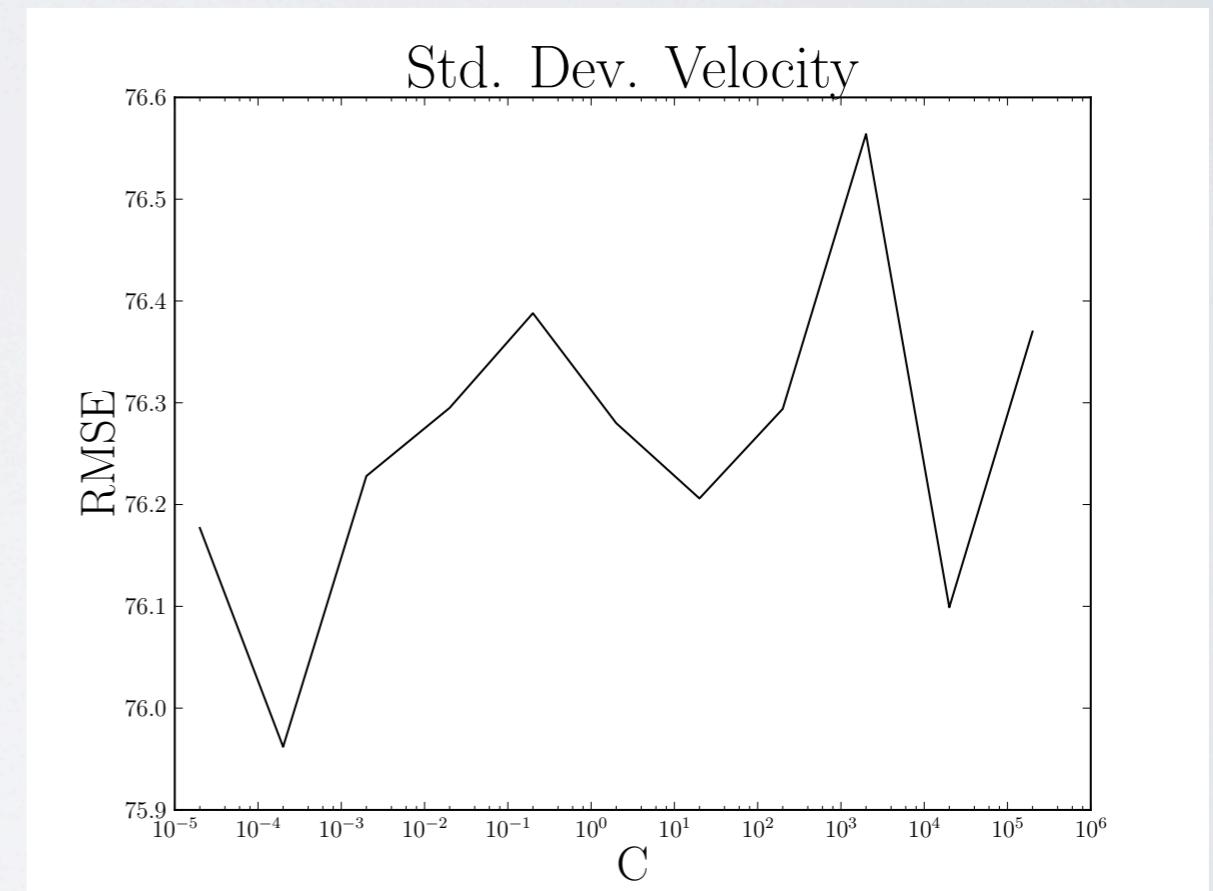
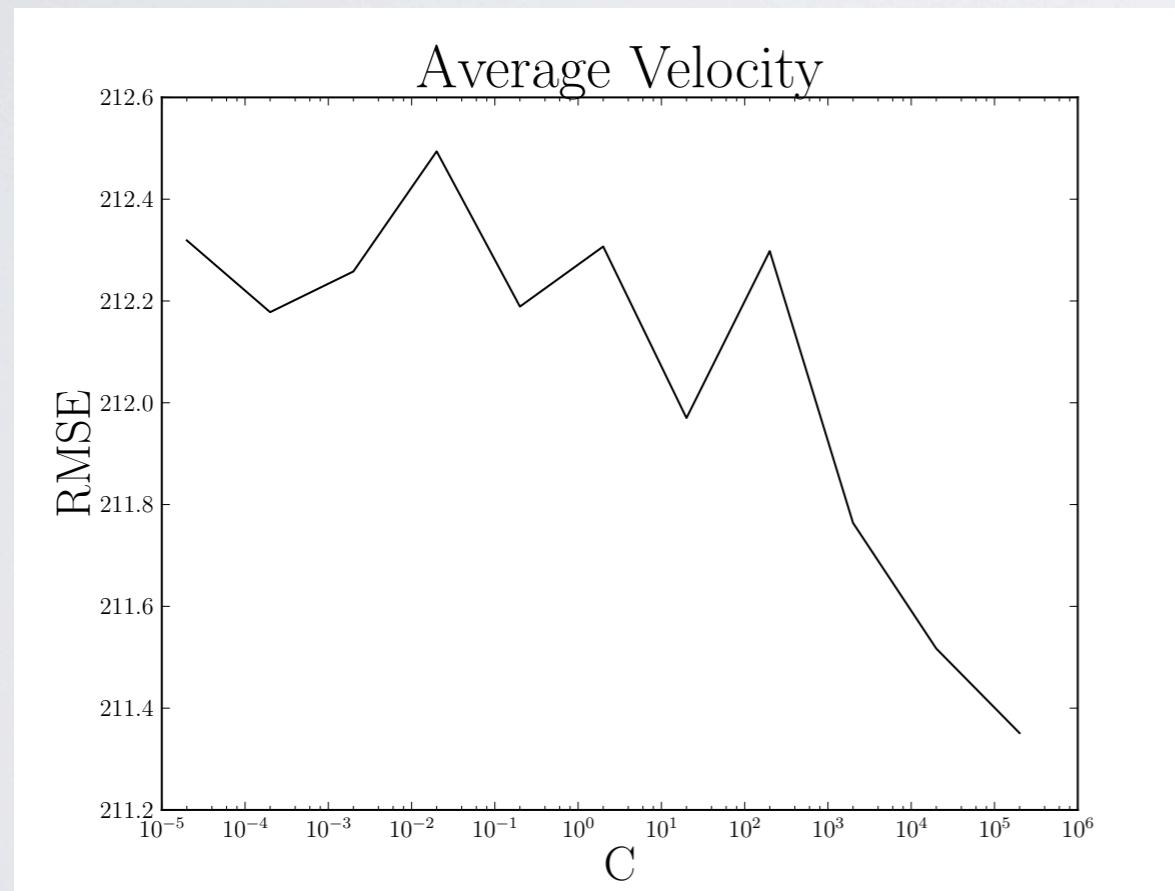


# PRELIMINARY RESULTS

- I've made four datasets:
  - tiny (10 samples)
  - small (100)
  - medium (1000)
  - large (10000)
- Today I'll present results for the medium dataset.

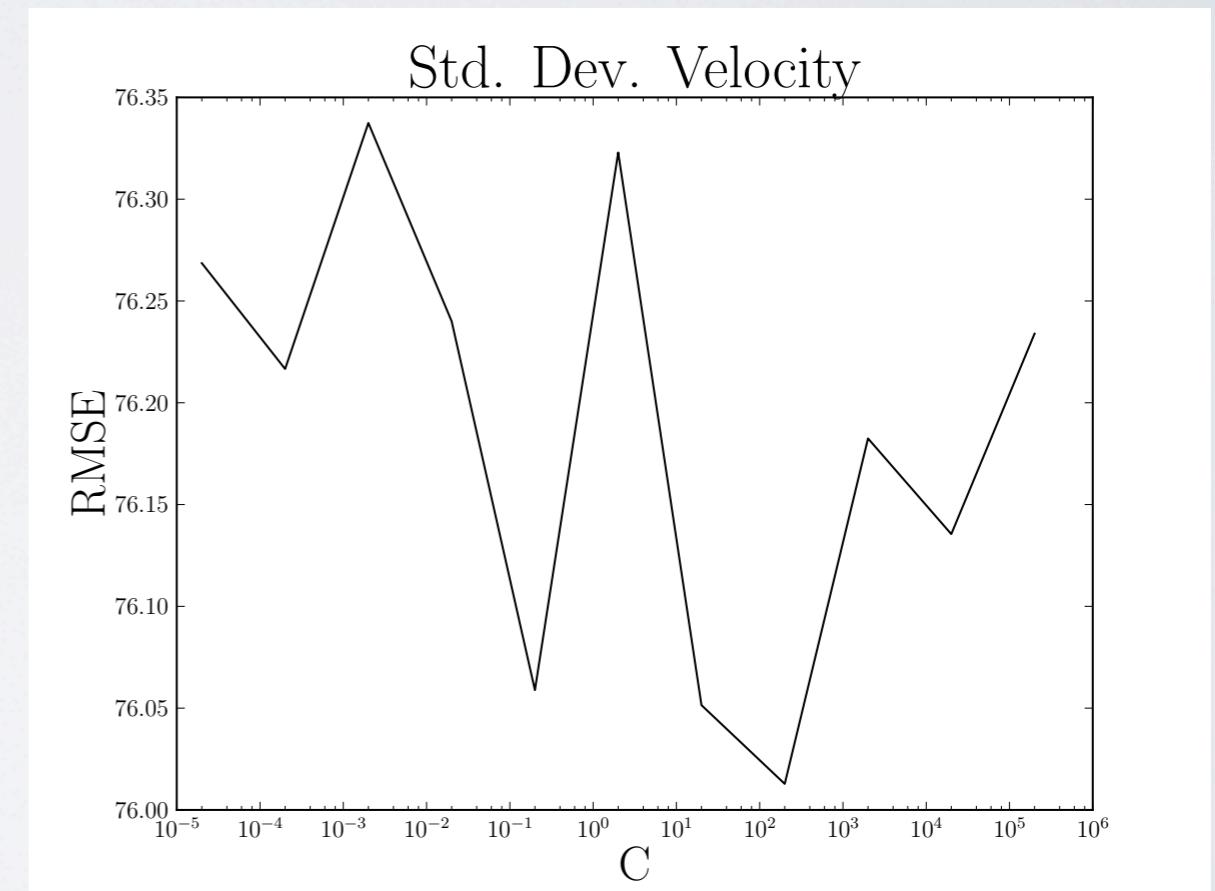
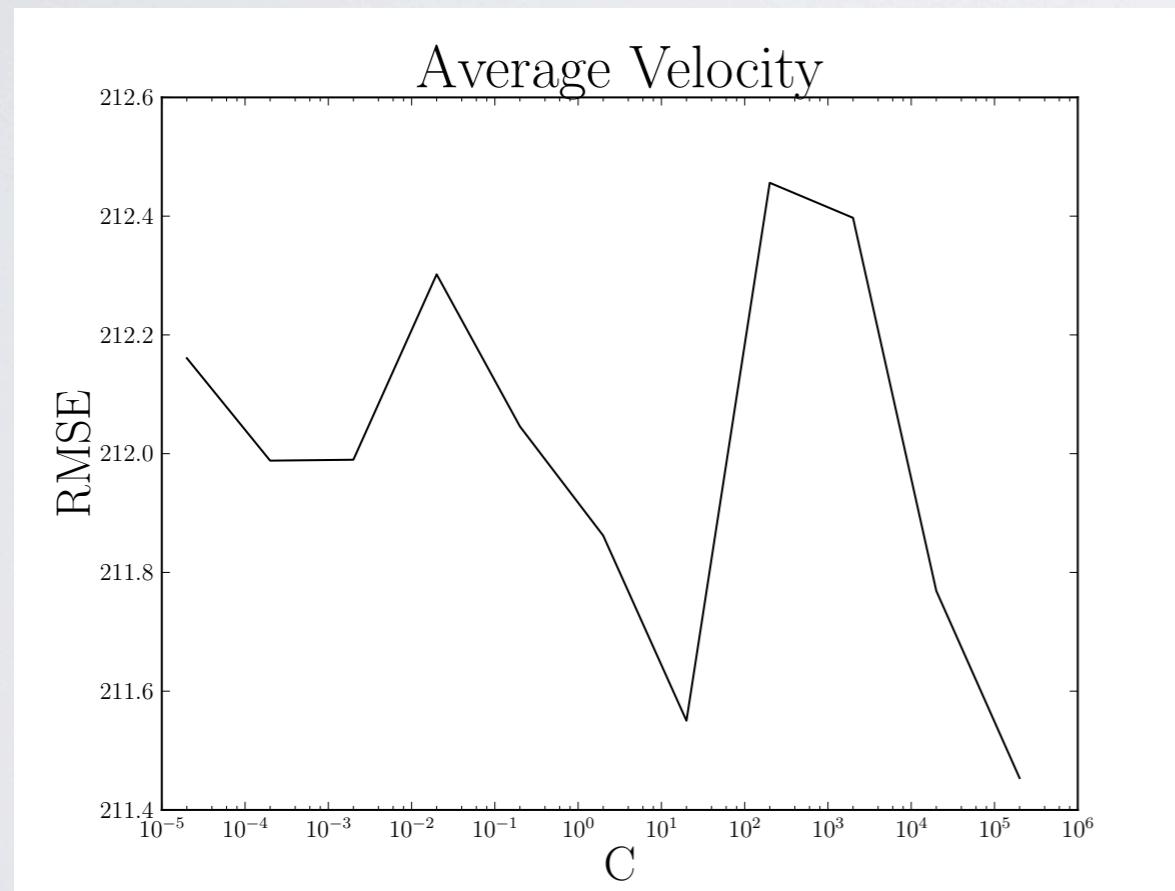
# PRELIMINARY RESULTS

- LI-normalized data, Linear Kernel, 5-fold cross-validation RMSE errors:



# PRELIMINARY RESULTS

- L2-normalized data, Linear Kernel, 5-fold cross-validation RMSE errors:



# PRELIMINARY RESULTS

- The errors using a degree-2 polynomial kernel and a gaussian kernel (with gamma from 2e-15 to 2e-1) are extremely similar. There is no clear advantage to using one kernel over the others.
- Presumably there is simply too little data at this point, so I will report back in more detail when I have results for the 10,000 - sample dataset.