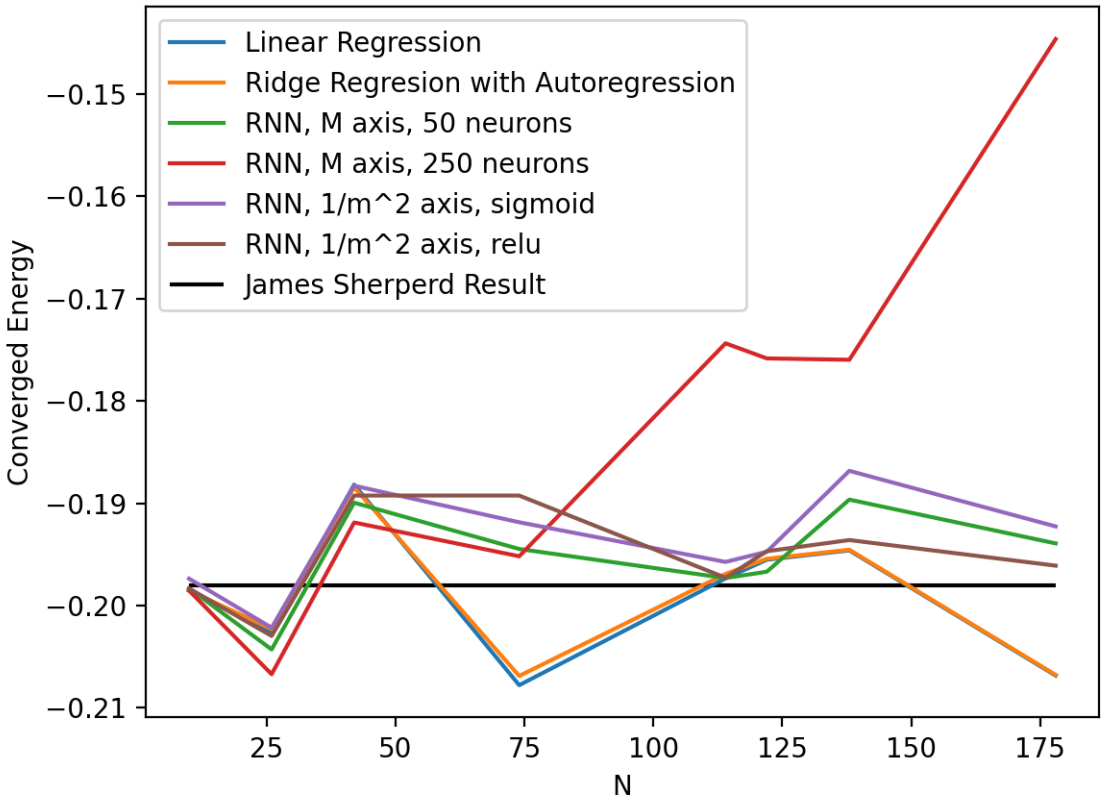

1. Linear Regression: Y data only using sequential formatting, Normalize and Set Intercept set to True. All data is training data for extrapolation. Extrapolated to a convergence threshold of 1e-6.

2. Ridge Regression with Autoregression: Y data only using sequential formatting. Hyperparameters found without using autoregression, Normalize set to True for extrapolation but alpha and solver are values found from hyperparameter tuning. All data is training data for extrapolation. Extrapolation found using autoregression to a threshold of 1e-6

3. RNN, M axis, 50 neurons: RNN found using traditional data formatting using M as the x data and correlation energy as the y data. RNN contained two feedforward hidden layers with 25 neurons each followed by three recurrent layer with 50 neurons each. Relu activation function for feedforward layers, sigmoid activation function for recurernt layers. Epochs and validation split varied to get optimal result. Extrapola-

tion was done by predicting the points at a variety of M values and taking the largest M value (9000) to be the converged energy.

4. RNN, M axis, 250 neurons: RNN with same architecture as previous point by each feedforward layer had 125 neurons and each RNN layer had 250 neurons. Epochs was set to 500 and the validation split was set to 0. Training and extrapolation performed same as previous point.

5. RNN, $1/m^2$ axis, sigmoid. Same architecture and parameters as point 3. Instead of the x data being M, it is transformed to be $1/(M/N)^2$. Training and extrapolation was the same as point 3, but with scaled x data.

6. RNN, $1/m^2$ axis, relu. Same as previous point but the activation function on the recurrent layers was set to relu.

7. James Shepherd Result: the converged value of the correlation energy

Notes:

1. The recurrent neural networks seem to consistently predict the converged correlation energies that are higher than what the data seems to suggest. This probably makes them a bad choice going forward since couple cluster already picks a higher energy than GFMC.

2. The recurrent neural network with 250 neurons definitely seems to be overfitting the data, especially at higher values of N, which correspond the the smaller data sets. After this run I switched the number of neurons to 50 which seemed to perform a better fit.

3. Using recurrent neural networks and traditional training data formatting seems to produce a smoother line than the regression and sequential formatting method, though none of the lines seems to display a consistent pattern.

Further Works:

1. It is possible that with 50 neurons the RNN is still overfitting given the size of the data sets. A full hyperparameter search should be done to find the best number of neurons, but this will likely need to be done on ICER.