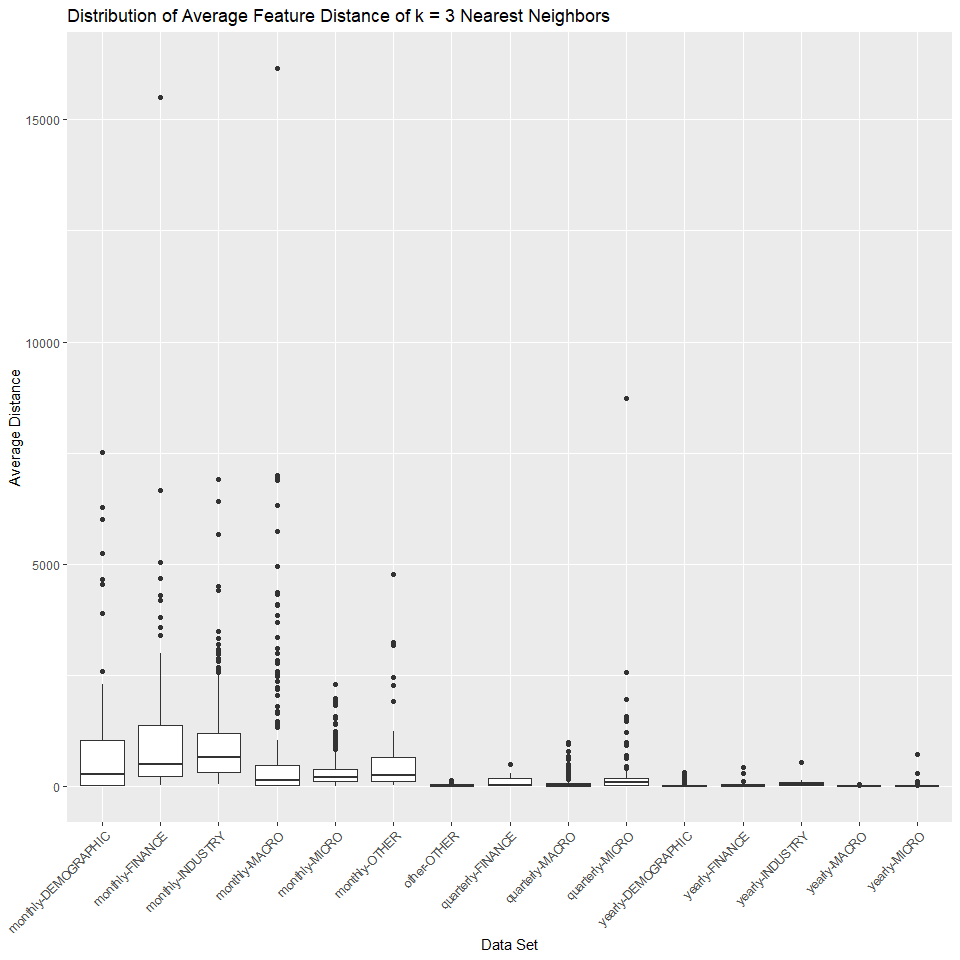
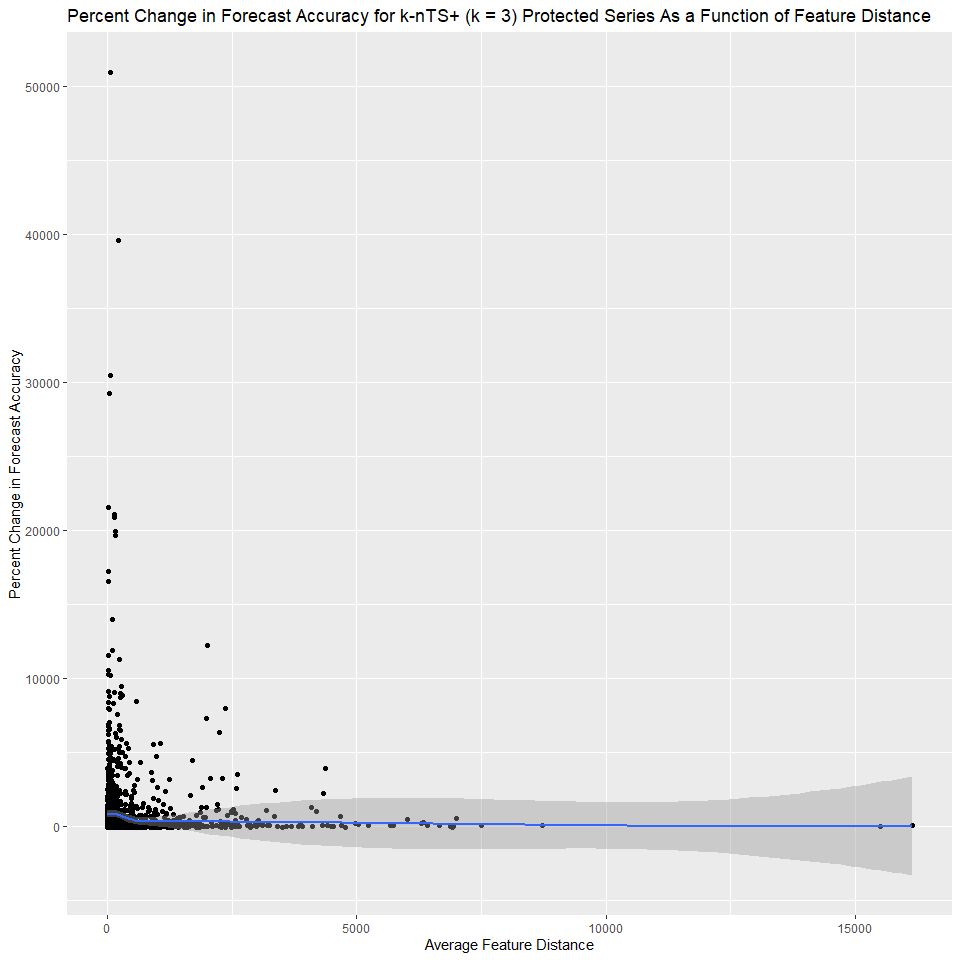
Let *FeatureDistance* denotea measure of how similar a neighbor is to a target time series based on the feature vectors and .

This is conveniently calculated using the formula for in the paper. I calculated for each series and averaged the values for the three nearest series, *i.e.*, the nearest neighbors based on features. The distribution of the average for each series in each data set is plotted below.

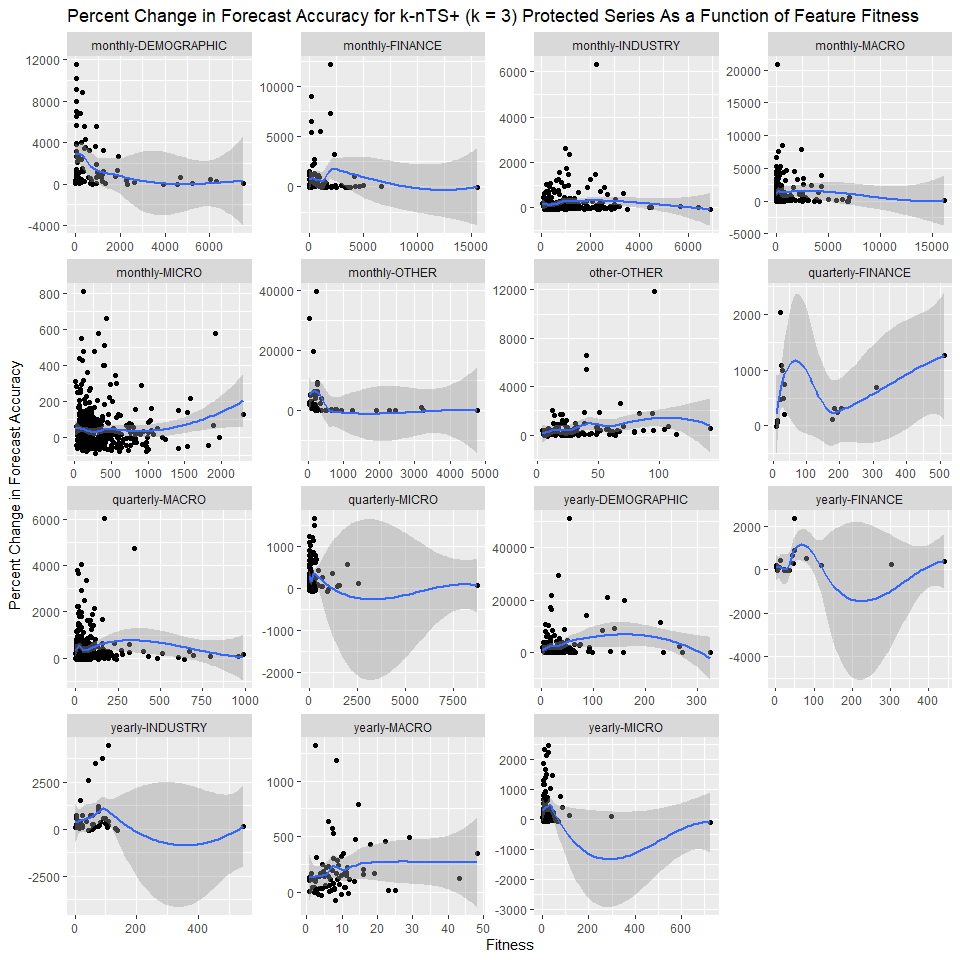


The average feature distance for the monthly data sets are larger than the quarterly, yearly, and ‘other’ data sets.

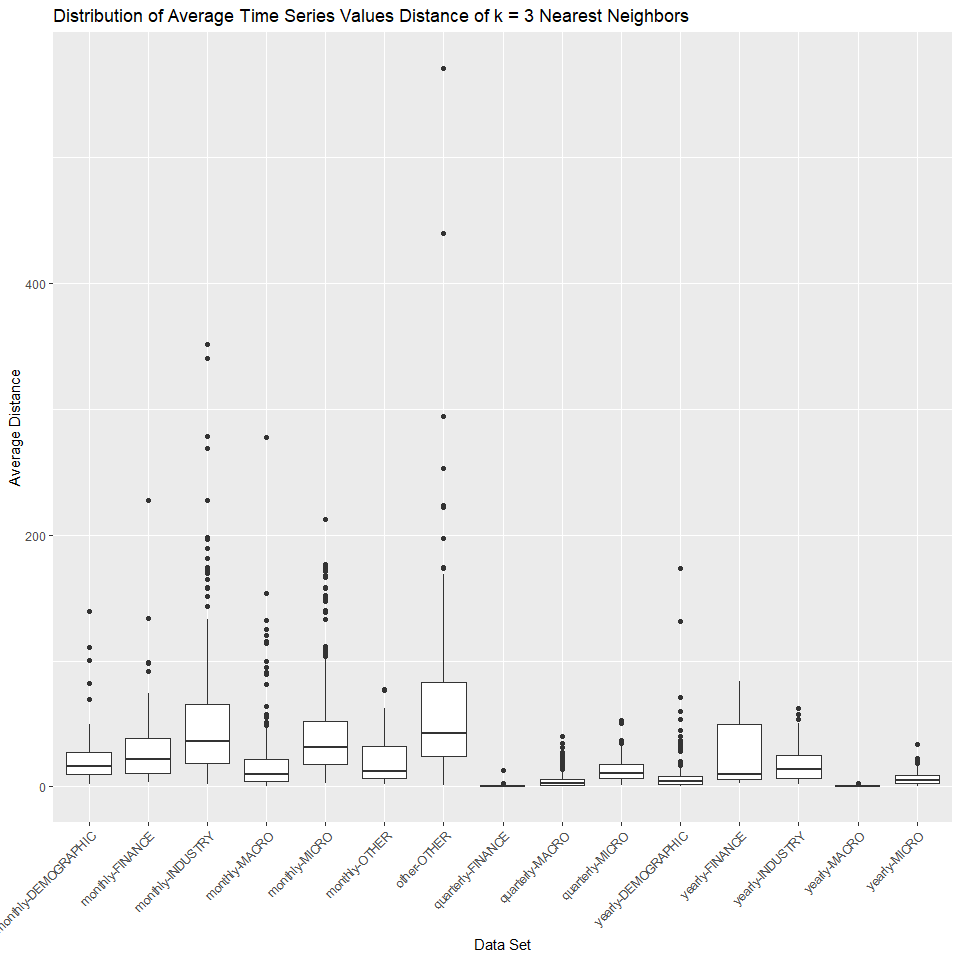
Next I plotted the average percent change in forecast accuracy across all our forecasting models against the average feature distance for all time series. Note that the regression line in this plot (and subsequent plots) is obtained from a Loess model. There doesn’t seem to be any significant correlation here. I expected to see the percent change in accuracy increase as the average feature distance increased. There are a lot of cases on the boundaries, *i.e.*, where there is little to no feature distance but large changes in accuracy, and vice versa.

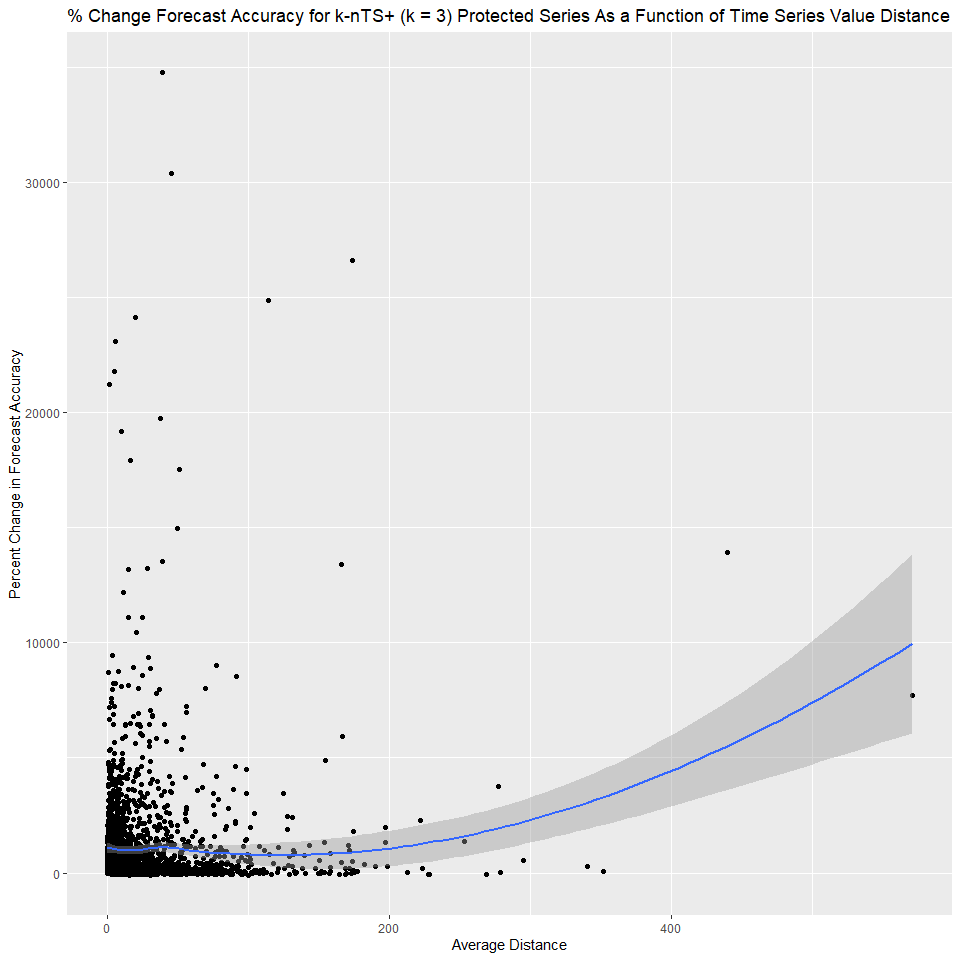


I recreated the above plot for each data set separately and got similar results.



My next thought was that maybe the similarity of the time series values matters once we isolate series with similar features for swapping. So, I repeated the above analysis but instead of the distance between feature vectors, I calculated the average distance between the time series themselves (conditional on being one of the three nearest neighbors based on the features).



The differences between frequencies are less clear cut here. The monthly series, which clearly had higher feature distances, have comparable value distances to some of the yearly series. Keep in mind this was calculated on log scale time series. Again, I expected a positive relationship between value distance and the percent change in forecast accuracy. In general, this doesn’t appear to be the case (the upward sloping line is from a Loess regression, and is due to the outliers on the far right side of the plot).

Again, recreating the above plot for each data set. There don’t really seem to be consistent patterns to pick up on here.

