

### Exercise: Predicting the Height of Waves

The problem concerns applying a PCA to predict the height of waves in the ocean.

- a) Split the data into a training set and a testing set. The only thing you need to turn in is a comment on what fraction of the total you used for each of the two sets.
- b) For your training data set, plot the relative error of the PCA as a function of  $k$ .
- c) The objective is to predict the wave height (WTMP) accurately. Assuming the PCA uses  $k = 1$ , plot the predicted WTMP values versus the actual values from the testing set. Note you should use the adjusted values here (as well as in what follows). Also, what is the MSME in this case?
- d) Redo part (c) for  $k = 2, 3, 4, 5, 6, 7$ .
- e) What is the recommended value for  $k$ ? Make sure to explain why. Also, does the plot in part (b) contradict any of your results from parts (c) and (d)?