Alice Ren

Data I

October 12, 2017

1. The two records match pretty closely. The automatic sampling temperature record has a greater range of values and variability on a given day, but the manual sampling temperature appears to follow day to day variations seen in the automatic sampling data. The plot for 2005 to 2017 is rather difficult to read, so a random year, 2013, is plotted as well. The light blue is the manual sampling bottom temperature while the red is the manual sampling surface temperature. The black is the automatic sampling temperature record.

The automatic temperature sampling takes measurements throughout the day. The manual shore station sampling occurs once per day. The time of day of sampling has varied over time, starting in the early part of the record around 8 AM but changing more recently to around noon. The time of sampling has been recorded concurrent to the temperature measurement since 1990, though sometime it is missing or is flagged as potentially erroneous. The shore sampling program has data beginning August 22, 1916, and data is available online (as of October 11, 2017) though October 31, 2015. The automatic shore sampling program has data from 2005 through 2017 available on the SCCOOS thredds server.

Though the prompt does not address the depth of measurement, I am going to average the manual sampling surface and bottom temperature and compare that value to the automatic sampling record. According to Melissa Carter (personal communication), the automatic sampling sensor is located around 2 meters depth. This is in contrast to the surface temperature at less than 0.5 m, and the bottom temperature at 4-5 m depth. (The netcdf file for SCCOOS displays the “depth” is 5m, which I take to mean the depth of the water at Scripps Pier – this is probably an erroneous value in the netcdf file. The readme file for the manual data says the bottom measurement is around 5m depth. The protocol for manual data readings takes water 1 m from the bottom for the bottom temperature. Thus, I feel that the bottom temperature being from 4-5 m is a reasonable assumption). A quick scan of the pressure data from the automatic sensor confirms that the automatic sensor is in the middle of the water column, though I did not take the time to find the average pressure (and thus average depth). For the remaining questions, I am using an average of the surface and bottom manual temperatures as the manual temperature record.

2. The automatic mean is 17.6994 +/- 0.0027 °C. The manual mean is 17.7861 +/- 0.0448 °C. The automatic mean’s error is smaller because N is over one million data points (1,069,188). From the means and their errors, the mean temperature is not consistent.

3. The standard deviation of the automatic data is 2.78 °C while the standard deviation of the manual data is 2.61 °C. The subsampled automatic data using the nearest neighbor to the time of collection of the manual data had a much larger standard deviation of 3.31 °C. The entirety of the automatic data actually had a better standard deviation than the subsampled compared to the manual, so it seems fine to use the entire automated record. I would say that the standard deviations are consistent because they differ by less than 0.2 °C, which is twice the measurement precision of 0.1 °C.

4. The Gaussian distribution is plotted with matlab, and the pdf is written in the code. Sketches of analytical solutions of the uniform and coin toss/bimodal distribution are below.

5. Plots attached in code.

6. The largest difference was greater than my test D statistic, which indicates I should reject the null hypothesis that both samples come from the same distribution. The K-S test suggests two different distributions. However, I should try a histogram with a smaller bin size to see if that affects D.

7. The manual temperature record was biased higher than the automatic temperature record from 2005-2015. Since the time of day collection was consistently at a certain time which may not be representative of the day’s average temperature, I would investigate the diurnal cycle from the automatic temperature record to rule out that the bias towards higher temperatures came from the sampling time. Another cause of bias could be from methodology. Melissa Carter’s explanation of the warming of the sample due to a warm PVC bucket was intriguing. Such procedural habits could create bias such as the one observed in this exercise. Since the variance is similar in the manual and automatic record, it is possible that an offset could be used, after it is determined whether the automatic record or the manual record is closer to the truth. It appears that the automatic sampling method has fewer user error opportunities, assuming that the sensor is cleaned and calibrated regularly. For temperature, both the automatic and manual sampling methods appear valuable. The manual records extend farther back in time, and if careful data quality measures are taken, the record could be extremely useful for long term, decadal trends. The manual record can also provide insight to when the automatic sampling is not working well. However, it appears that the manual sampling could be improved to understand its accuracy, especially with respect to the automatic sampling method.