Problems Week 2

Due Monday, October 19, 2020

As we've noted in class, data sets have been collected at the end of the pier for 104 years, and at present SIO collects continuous automated measurements and once-per-day manual measurements. One of the challenges for a time series of this duration is to decide whether continuation of the volunteer-intensive manual effort is fully justified, as Melissa Carter, the Shore Stations Program Manager, will likely explain to us. The automated data are available for download here:

http://sccoos.org/thredds/catalog/autoss/catalog.html,

The newest version of the manual data for surface temperature are here:

https://shorestations.ucsd.edu/wp-content/uploads/sites/366/2020/03/RazTS_SURF.csv

and are described in the recent paper (Rasmussen et al, 2020) available from Canvas. You can also find the raw data and other information at the Shore Stations web site (https://shorestations.ucsd.edu/data-sio/ under the "Adjusted Time Series" tab.) The manual data are CSV (comma-separated-variable) files, often seen for spreadsheets. You can read these in Matlab with "csvread" or "dlmread" and in python with "numpy.genfromtxt" or "pandas.read_csv". I had some trouble with the CSV file provided on the data set, so will post my version (with commas replacing all other punctuation) on Canvas.

1. **Visual evaluation.** Plot the two records. Are there obvious discrepancies between them? What time of day are the samples collected? What years are covered? What has been done to adjust the manual data for time of day?

2. **Means.** Evaluate whether the mean temperature is consistent within error bars.

3. **Variance.** Evaluate whether the variance or standard deviations are consistent. Do you obtain different results if you subsampled the automated data at the times of manual measurements, or if you use all of the automated data?

4. **Theoretical pdfs.** For the automated data, given the observed mean and variance, what would the pdfs look like if the data were drawn from (a) a Gaussian distribution, (b) a uniform distribution, or (c) a bimodal (coin-toss) distribution? You can do this analytically or numerically (e.g. using Matlab). Bonus points if you can show both methods.

5. **Empirical probability density functions.** Plot the pdfs for the automated and manual records.

6. **Compare pdfs.** Evaluate (quantitatively) whether the observed pdfs are comparable within error bars.

7. **Summary.** On the basis of your evaluations, what is your assessment of the value of the manual data relative to the continuous data?

As we'll see on the pier tour, there are some subtleties to the pier data sets, particularly for the automated conductivity sensor (which is used to derive salinity—not part of this problem set). Standards for maintaining the automated system have improved since 2012.