STAT 248 PROJECT PROPOSAL

SARA STOUDT

1. Data

I have data from collaboration with the San Francisco Estuary Institute for a different project that I did last summer. The data is monthly water nutrient measurements from multiple stations spread across the San Francisco Bay Delta.

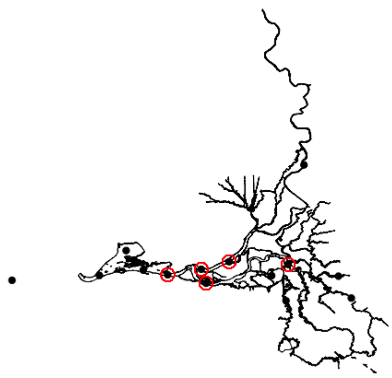

Figure 1. All stations marked with black dots. I'm narrowing my focus to the stations marked in red as they have more continual records and are in the part of the Bay Delta system where we would expect the largest correlations based on previous work.

I will narrow my focus to 5 nutrient and water quality measurements:

- chlorophyl
- dissolved oxygen
- pheophytin a
- temperature
- salinity

2. Motivation

Long-term time series of water quality measurements spanning multiple locations in the San Francisco Estuary are available. The challenge is to leverage this data to learn how nutrient concentrations vary across space and time in the Delta, what the major drivers of this variability are, and if any mechanistic insights can be gained by exploring the relationships between various variables collected.

Questions to answer and insights for scientists that can be gleaned:

(1) Within a station, are different nutrients significantly correlated? This knowledge can help scientists generate hypotheses of water quality mechanisms at play locally.
(2) Across stations, are the same nutrients significantly correlated? This knowledge can help scientists assess connectivity and predict where water quality damage could spread.
(3) Across stations, are different nutrients significantly correlated? This knowledge can help scientists generate hypotheses of more bigger picture water quality mechanisms at play in the whole system.
(4) Can we strengthen our claims about anything we find correlation-wise into causation? This knowledge would help strengthen scientists cases when recommending water quality management decisions.
(5) Can we quantify the efficiency of each station in terms of the information about the system that it provides? This knowledge could help with future measurement network planning to ensure the most effective allocation of resources while maximizing the information about water quality across the Delta.

3. Methods

(1) Pick a time interval with near complete records. Use hotdecking to fill in the few remaining missing values.
(2) Pick out the stationary part of each series to work with (subtract loess curve). Use variance stabilization transformations as needed.
(3) Assess significant cross correlations. Simulate null distribution using shuffling in blocks, and estimate uncertainty using the jack-knife.
    - Within a station: cross correlation between nutrients
    - Across stations: cross correlation of same nutrient
    - Across stations: cross correlation of different nutrients
(4) Work in the frequency domain to support the above step's work in the time domain using coherence.
(5) Try to strengthen to causality, testing significant cross correlations from above using mutual information. Again simulate null distribution using shuffling in blocks, and estimate uncertainty using the jack-knife.

(6) Test significant partial coherence given a station in between two stations to quantify the efficiency of each station's information. Given a station in between, is there still a significant correlation between the same or two different nutrients? Simulate null distribution using shuffling in blocks. Estimate uncertainty using the jack-knife.