

Predicting ocean toxin levels in California coastal region

Introduction

While algae are important players in our ecosystem and the marine food web, under certain conditions, some algae can grow rapidly and release large amount of toxins that are harmful to human and other marine animals. Such biological events are known as harmful algal blooms. There are many types of harmful algal blooms; one of them is of particular interest to pacific coastal states of the U.S. That is *Pseudo-nitzschia* bloom, which produces the toxin, domoic acid. Domoic acid is a neurotoxin that can cause serious neurological symptoms, and sometimes death, if ingested. It is not only dangerous to swimmers and beachgoers, but it can also accumulate in seafood, especially shellfish such as crabs and clams. Therefore, high domoic acid level in coastal waters is not only a [public health concern, but also an economic risk](#). For example, the harmful algal bloom in 2015 caused [closures and delays](#) of dungeness crab, rock crab season, which costed [millions of dollars](#) to the fishing/seafood industry.

Problem

Physical, chemical, and biological conditions such as water temperature, nutrients in the water, seasonality and other algae in the water may contribute to the formation of harmful algal blooms and the release of toxins. This project aims to use these data collected by federal and state governments to predict toxin levels coastal waters.

Clients

Because of public health and economic implications of toxin levels in coastal waters, many entities will be interested in the prediction of those levels. They include government agencies in charge of public health, environmental protection and fishing regulation. The fishing and aquaculture industry, which is worth billions of dollars, will also be heavily interested in the prediction. Downstream industries such as restaurants and packaged seafood industry might also be interested.

Data

Though there are data available for all states on the pacific coast of the U.S., this study is going to focus on California. Southern California Coastal Ocean Observing System has been monitoring domoic acid levels and accompanying physical, chemical, and biological data since 2008. Weekly data of 8 stations from San Francisco to San Diego can be downloaded via their [website](#) in CSV format. This will be the main dataset. Other oceanography data will also be added. They include daily ocean upwelling indexes provided by Pacific Fisheries Environmental Laboratory. This [dataset](#) include daily data of 15 stations for the past 50 years, though only 2 stations and the past 10 years are relevant to our main dataset. The Oceanic Niño Index (ONI), which monitors climate variability in the ocean, will also be added. This [dataset](#), provided by the National Oceanic and Atmospheric Administration is available in tabular format, as 3-month averages.

Analysis plan

Data of different sources will be downloaded. Relevant data will be extracted, and merged together, and cleaned. Exploratory data analysis will be conducted to see which factors might have impact on toxin levels. Regression models can be built to predict the concentration of toxin levels. Alternatively, toxin concentrations can be categorized as dangerous and not dangerous, or several risk levels. Classification models can then be built to predict the risk categories. Same day prediction will be tried first, then future prediction (1 or 2 week in advance) will also be attempted.

Deliverables

Deliverables will include cleaned data, all codes in notebooks and a report or presentation.