Development and application of open source tools for the analysis of long-term trends in coastal water quality

Dr. Marcus W. Beck

February 12, 2018

## Overview

Open science represents both a philosophy and a set of tools that can be leveraged for more effective scientific analysis. At the core of the open science movement is the concept that research should be reproducible and transparent, in addition to having long-term provenance through effective modes of data preservation and sharing. Open science is critically important for fields of research that both collect and utilize monitoring data from aquatic ecosystems. Techniques from open science can help technicians, researchers, and managers address key challenges of these data, including extraction of information from multiple sources.

Several programs across the country have been collecting water quality data that are extensive in space and time. Well-known examples of long-term monitoring programs exist in prominent estuaries, such as Tampa Bay, Chesapeake Bay, and San Francisco Bay. The National Estuarine Research Reserve System (NERRS) has also collected continuous data at 29 smaller estuaries since 1995. In southern California, the Bight regional monitoring program has collected coastal data at five-year intervals since 1994. Together, these data represent a valuable resource that can be leveraged to evaluate the status and trends in water quality at multiple spatial scales.

Effective environmental management can depend on the definition and description of past trends to inform decision-making. To guide this effort, long-term monitoring data are critically important to identify environmental changes relative to anthropogenic and natural drivers. However, the ability to quantitatively link drivers of change with environmental trends is complicated by lack of analysis tools to effectively evaluate existing datasets. Specific challenges for coastal environments include:

* Significant variation in data products to evaluate trends, including different time periods, spatial coverage, and measured parameters,
* No consensus on which analysis tools best inform decision-making through description of trends,
* Many existing analysis products are not openly accessible, limiting transparency and reproducibility in the scientific community.

Although these challenges represent broad needs, the proposed research will focus on two specific topics that have relevance for water quality issues in southern California. The first focus area will develop a unified method for trend analysis of primary production across coastal environments. The second focus area will develop a data synthesis and analysis method to link coastal restoration projects with environmental changes using a risk-assessment framework.

## Benefits to SCCWRP and its member agencies

Both focus areas will adopt an open science workflow to evaluate water quality from existing monitoring programs. This approach will inform the broader management and scientific community on methods for data synthesis and trend analysis to inform decision-making. Environmental data from southern California will be contrasted with additional state and national datasets to identify common themes or specific lines of evidence that can improve coastal management. Specific products will include a manuscript and a software package created for each focus area. All datasets will be made accessible in an open data repository.

## Focus Area 1: Interbay comparisons of drivers of primary production using generalized additive models

Recent applications of statistical methods to describe water quality changes have shown promise in coastal environments. The need to understand drivers of primary production is particularly important near estuaries or open embayments that act as mixing areas of pollutants from the watershed. Generalized additive models (GAMs) have been used to describe eutrophication endpoints in tidal waters and are particularly appealing because they are less computationally intensive than alternative methods and provide accessible estimates of model uncertainty. The Chesapeake Bay Program has begun development of an open-source software program to identify water quality trends using GAMs. However, this program is limited in scope and has not been applied to other datasets outside of Chesapeake Bay. Application of a common method to multiple locations could prove beneficial for informing our understanding of geographic differences in drivers of coastal eutrophication. Specific questions include:

* Can a unified GAMs method be used to distinguish the relative impacts of climate, physical, and chemical drivers in estuarine and coastal primary production?
* Can these methods be used to quantify differences over time and between systems as related to controlling factors?
* What challenges are encountered that identify limits of the modelling approach?

We will focus on routine monitoring data of chlorophyll from San Pedro Bay at the outflow the Los Angeles River. These data will also be compared with long-term time series from San Francisco Bay, Chesapeake Bay, and Tampa Bay. The proposed research will develop a hypothesis-testing framework for GAMs to identify varying contributions of climate, physical, and chemical drivers. We will quantify the variance in production that is explained by these different drivers to develop conclusions on the importance of each controlling factor. An expected outcome is an analysis method of trends that can identify which factors are relevant for different regions that can help prioritize management actions.

## Focus Area 2: Use of open science to inform coastal restoration projects

Habitat and water quality restoration projects are commonly used to enhance coastal resources or mitigate negative impacts of water quality stressors. Significant financial resources have been contributed for restoration projects, yet much less attention has focused on evaluating the outcomes beyond site-specific assessments. An understanding of the large-scale effects in both space and time on changes in water quality is critical to quantify the return on investments in restoration projects. Initial work on this project has focused on evaluating multiple datasets in the Tampa Bay area to identify:

1. What types of restoration projects produce the greatest improvements in water quality, and
2. Which time frames and synergistic effects of projects are most relevant for having the largest perceived benefits?

Changes in chlorophyll concentrations as a proxy of eutrophication were used to assign a probabilistic expectation of water quality changes from investments in restoration activities. A unique approach was developed to synthesize datasets that maximized identification of a significant change in water quality from different types of restoration activities. This project was implemented specifically with a diverse group of collaborators as part of an open science training workshop at UC Santa Barbara in July 2017. Methods developed from this project will be extended to evaluate cumulative effects of restoration projects on water quality in coastal zones of southern California. An expected outcome is the ability to quantify the relative tradeoffs in investment associated with different types of restoration projects. The analysis will also aid in identifying how many projects and approximate time frames within which an improvement in water quality is expected.