A Program Suite for Visualizing and Interpreting Four-Dimensional Environmental Monitoring Data, and its Application to Locating Pollution Sources

Scott Ritter
Civil Engineering and Environmental Engineering

Author

Scott Ritter, while rowing on Newport Bay as a member of the UCI Rowing Team, became interested in the Bay's water quality and the sources of pollution that have compromised it. Through his work with Dr. Stanley Grant, Scott has developed a computer program for visualizing and interpreting environmental data, a potentially valuable tool in maintaining and restoring the quality of water in the Bay. He describes research as "a priceless introduction to the world of science and academia," and has particularly enjoyed the vicissitudes of the research process—the fluctuations between unavoidable failures and unimaginable successes. Since Fall 2005, Scott has been pursuing a graduate degree at Stanford University.

Key Terms

- Estuaries
- Fecal Indicator Bacteria
- Harbors
- Multivariate Data Visualization
- Tidal Saltwater Marshes

Abstract

Water quality monitoring data are often four dimensional—including three spatial dimensions and time. The multidimensional nature of water quality monitoring data poses significant interpretational challenges to researchers and managers interested in identifying and mitigating sources of water pollution. These interpretational challenges are particularly acute in tidal embayments (e.g., coastal wetlands and marinas), because the direction of pollution transport (inland or coastward) constantly changes with the tides. To address these challenges, a generalized visualization computer program was developed and used to analyze 4,132 measurements of fecal pollution in Newport Bay, a regionally important tidal embayment in southern California. The results of this analysis indicate that most fecal pollution in the water column of Newport Bay originates in runoff from local storm drains and creeks. Once contaminated runoff flows into Newport Bay, fecal pollution associated with the runoff is transported laterally by the tides. The computational tools developed and tested in this study generate user-friendly graphics that enable a rapid interpretation of large multivariate datasets. Furthermore, the computer programs are implemented in a powerful graphics program (Igor, Wavemetrics) and can be easily used to visualize spatial and temporal pollutant trends in a wide variety of environmental settings.

Faculty Mentor

When it comes to coastal water quality, Southern California is a "perfect storm," in its juxtaposition of an urban megalopolis and a world-class coastline that serves as a recreational destination for literally millions of beach visitors each year. Scott's work is significant in providing a way to visualize, and thus better understand, the role that critical transition zones, such as Newport Bay, play in buffering coastal water quality against the impacts of polluted surface water runoff from urban landscapes.

Stanley B. Grant
Henry Samueli School of Engineering

Introduction

Environmental monitoring data—collected to study water quality, tidal mixing, and other environmental parameters—are often large and multidimensional. This large, multidimensional nature poses significant interpretational challenges to researchers and managers interested in identifying and mitigating sources of water pollution. These interpretational challenges are particularly acute in tidal embayments (e.g., coastal wetlands and marinas) because the direction of pollution transport (inland or coastward) is constantly changing with the tides. Oftentimes, temporal and spatial visualizations of environmental monitoring data are necessary for a better understanding of complex environments such as tidal embayments (Takahashi et al., 2000).

To address the need for temporal and spatial visualizations of environmental monitoring data, several tools have been developed (Takahashi et al., 2000; Grunfeld, 2005). These tools include freeware visualization programs (e.g., XmdvTool and XGobi/GGobi) that create scatterplot, star glyph, parallel coordinate, and dimension stacking figures (Grunfeld, 2005), and a visualization system that uses two-step linear interpolation to create useful contour plots for sparsely collected data—collected over irregular intervals over irregular grids (Takahashi et al., 2000). These tools can be used to produce useful representations of multidimensional spatial and temporal data. Nevertheless, they have respective drawbacks in that separate scatterplot, star glyph, parallel coordinate, and dimension stacking figures can be challenging to interpret, and smoothing sparse data through two-step linear interpolation can produce false representations of temporal and spatial variability.

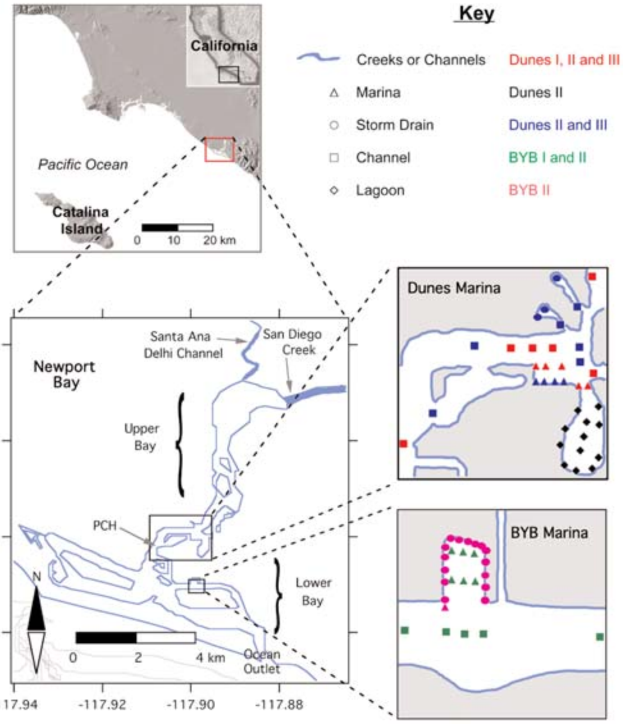

Figure 1
Maps of Newport Bay, showing the Dunes Marina and Balboa Yacht Basin (BYB) study sites

To overcome these drawbacks to visualizing and interpreting environmental monitoring data, we developed a suite of visualization programs, which we used to analyze 4,132 measurements of fecal pollution in Newport Bay, a regionally important tidal embayment in southern California. Newport Bay (hereafter referred to as the Bay) is the second largest estuarine embayment in southern California (Figure 1). The Bay provides a critical natural habitat for terrestrial and aquatic species, and is a spawning and nursery habitat for commercial and non-commercial fish species (California Regional Water Quality Control Board, 1999). The lower portion of the Bay (Lower Bay) is a regionally important recreational area and one of the largest pleasure craft harbors in the United States. The upper region (Upper Bay) is a state ecological reserve and provides refuge, foraging areas, and breeding grounds for a number of threatened and endangered species. The Pacific Coast Highway bridge (PCH in Figure 1) roughly demarcates the boundary between Upper Bay and Lower Bay.

This paper presents a suite of generalized visualization programs that reduce interpretational challenges by producing simple visualizations, and presents the program suite's application to locating pollution sources using environmental monitoring data collected from the Bay.

Methods and Materials

A suite of computer programs—an animation generation program (AGP) and a layout generation program (LGP)—was developed to help interpret complex data collected from water quality monitoring programs, specifically, to help locate pollutant sources in the Bay. Each program is a portable platform, developed using Igor Pro (programmable graphing software produced by Wavemetrics, Inc.), that processes specific parameters for different case studies. These programs are preliminary analysis tools, which can guide subsequent, more rigorous statistical analyses. The difference between them is that, after creating multiple graphs (one per animation frame), the AGP concatenates them into a movie file, whereas the LGP tiles, then appends color scale footnotes to produce a layout. The

AGP is best used for video presentations, whereas the LGP is best used for print presentations.

The purpose of the AGP is to rapidly translate complex data sets into simple animations. The AGP takes data—input by the user from a user-defined environmental variable—and automatically generates animation frames. These frames are then automatically combined to form animations that display complex multivariate data both spatially and temporally. Sample animations are shown as Figures S1–S3 in the Supporting Information, which is located at www.urop.uci.edu/framejournal2005supplement.html.

The LGP operates on principles similar to those of the AGP; but, instead of generating animations, it automatically appends a user-defined number of animation frames to layouts (windows containing graphs, tables, annotations, or pictures) for printing purposes (examples are shown in Figures 4 and 5, and in the Supporting Information, Figures S4–S7). Like the AGP, the LGP simplifies the visualization of complex data by combining multiple variables into one figure.

The AGP/LGP contours data for each temporal variable through one-step, bilinear interpolation between spatial—not temporal—points, while setting contour boundaries to the maximum and minimum data values and scanning each matrix element to convert erroneous zeros into "NaN" (not a number). A land feature façade, derived from traces of GPS data, is layered over the contour for geographical reference. A time-stride tide graph and color scales are appended.

The program suite was used to assess the contribution of marinas to fecal indicator bacteria impairment of the Bay. To achieve this objective, five separate field experiments were carried out at two different marinas: the Balboa Yacht Basin (BYB) and the Dunes Marina (Dunes). The two marinas differ with respect to their size (BYB has 174 boat slips, Dunes has 450 boat slips), shape and, presumably, tidal flushing characteristics (see Figure 2 for aerial photos of the BYB and Dunes marinas).

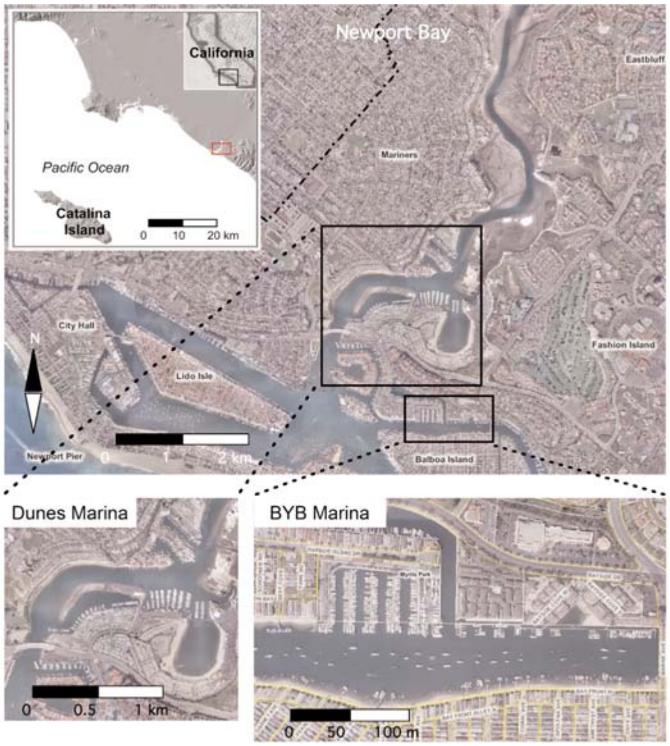

Figure 2
Aerial photos of Newport Bay, Dunes Marina, and BYB Marina

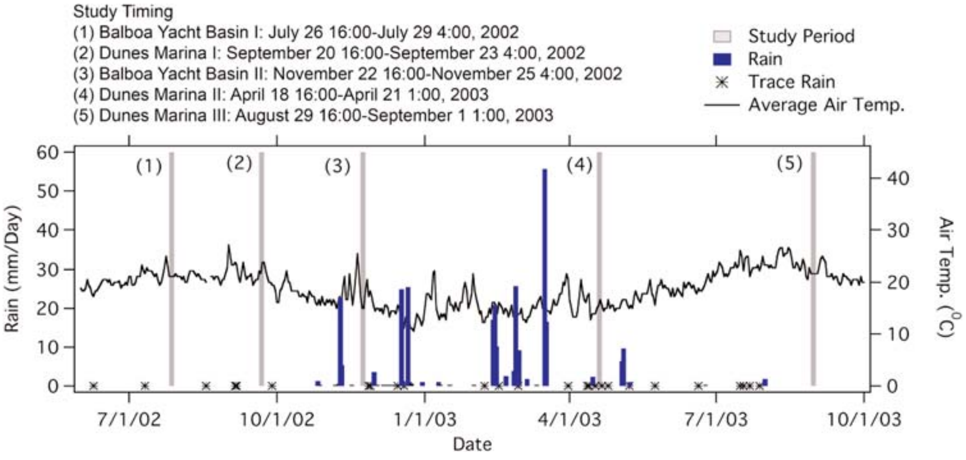

Figure 3
Timing and general features of the five marina studies

Figures 1 and 3 summarize key features of the experimental design employed in each of the five studies. Altogether, 4,132 water samples were collected from the BYB and Dunes marinas over a 14-month period, from July 2002 through September 2003. All five studies were conducted during dry weather periods; however, the elapsed time from the last storm varied considerably, from two days for the second Dunes study to 123 days for the first Dunes study (Figure 3). All five studies had similar experimental designs, although the number of samples collected and the spatial distribution of sampling sites varied.

Sampling sites were divided into four categories based on their location: marina samples, collected from sites located within the marina being studied; channel samples, collected from sites located within the channel adjacent to the marina; storm drain impacted samples, collected from sites in the marina or channels adjacent to storm drain outlets; and lagoon samples, collected during the second Dunes study from a lagoon near the Dunes marina. The location of sampling sites within the BYB and Dunes Marinas are indicated in Figure 1.

The sampling frequency and the number of sampling sites for each of the five studies were based on several objectives. To assess the reproducibility of our sample collection and analysis methods, duplicates were collected for 10% of all samples (refer to Jeong *et al.*, 2005 for the results of this objective). To capture tidal and diurnal (day/night) changes in the fecal indicator bacteria signal, water samples were collected every three hours, twenty four hours per day, for the duration of each 2.5 day study period. To assess vertical stratification of fecal indicator bacteria in the water column, paired water samples were collected from each site, one from the surface of the water column and another from 1 meter below the surface. For all practical purposes, fecal indicator bacteria pollution in Newport Bay is well mixed down to at least a depth of 1 m, thus eliminating the need for a vertical dimension in the analysis programs (refer to Jeong *et al.*, 2005 for the results of this objective). To maximize the possibility of detecting illicit discharges from vessels inside the marinas, studies were conducted from Friday afternoon through Monday morning, when boat usage was significant. Several of the studies were conducted during holiday periods. All samples were transported a short distance to UCI, where they were analyzed for fecal indicator bacteria—including total coliform (TC), Escherichia coli (EC, a major group of fecal coliform), and enterococci bacteria (ENT)—using defined substrate tests, known commercially as Colilert and Enterolert, implemented in a 97-well quantitray format. All water samples were also analyzed for salinity, turbidity and pH, using methods described in Reeves *et al.* (2004); details of the water column sampling, including the logic used in selecting the sampling sites, can be found in Jeong *et al.* (2005).

Results and Discussion

Figures 4 and 5 (and Figures S1–S3 in the Supporting Information), generated by the LGP and AGP respectively, were used to visualize site-to-site and study-to-study variability and correlations between fecal indicator bacteria and salinity for the Newport Bay field studies. The difficulty in data interpretation and communicating results—using individual box plots, time series, and contour plots—has been simplified through use of the AGP/LGP.

Figure 4, generated by the LGP, shows that the single-sample exceedences—instances when single water samples exceeded the California standards—and geometric means of fecal indicator bacteria are nearly the same across site categories (i.e., marina, channel, storm drain impacted, and lagoon). Exceptions to this rule include the second BYB study, for which samples collected adjacent to storm drains (particularly along the west wall of the marina) had elevated concentrations of TC and ENT (third row of color panels in Figure 4) and the second Dunes study in which samples collected at various channel stations and along the eastern shoreline of the lagoon had elevated concentrations of ENT (fourth row of color panels in Figure 4).

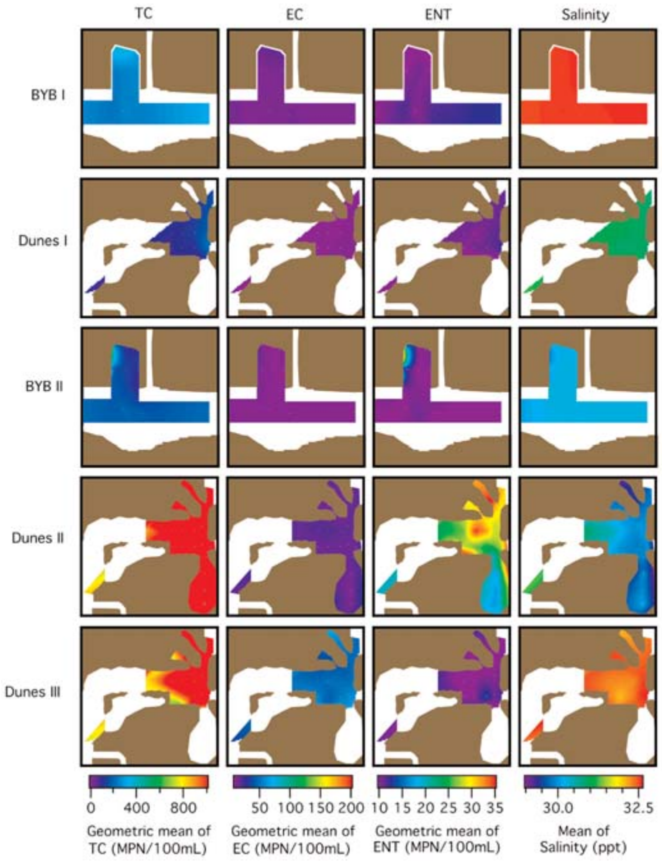

Figure 4

This figure, generated by the AGP/LGP program suite, depicts site-to-site and study-to-study variability of the geometric mean of fecal indicator bacteria (Most Probable Number per 100 mL) and mean salinity concentrations (parts per thousand) measured across the five Newport Bay studies. Relative color homogeneity within panels suggests little site-to-site variability, whereas color differences between panels suggest significant study-to-study variability.

Figure 4 also shows that while the concentration of fecal indicator bacteria generally exhibits relatively little site-to-site variability, significant study-to-study variability is evident. The highest single-sample exceedence rates and geometric means occurred during the second Dunes study, during which 10% of all samples exceeded the single sample standard for ENT (notice the high concentrations in the fourth row, third column of Figure 4). The second Dunes study also had the shortest antecedent dry period (2 days, Figure 3) and the lowest recorded salinity values (Figure 4). The collective observations from the LGP-generated Figure 4—low mean salinity concentrations paired with high geometric mean fecal indicator bacteria concentrations, most notably for TC and ENT during the Dunes II study—and the supporting data (see Supporting Information, Table S1) suggest a potential connection between surface water runoff (which has lower salinity than ocean water) and fecal indicator bacteria impairment in the Bay.

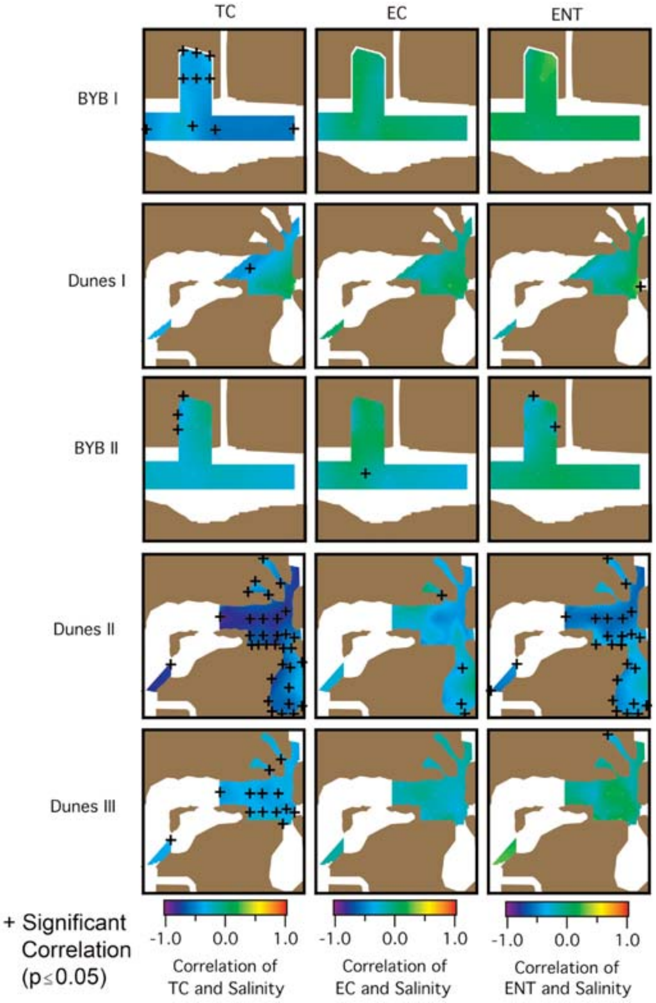

Figure 5

Spearman Rank correlations between fecal indicator bacteria and salinity across the five Newport Bay studies. The negative correlation between fecal indicator bacteria and salinity at many sites in both the BYB and Dunes field areas—most notably during studies with short antecedent dry-weather periods (see Figure 3)—is consistent with a runoff source for these organisms.

The above site-to-site and study-to-study observations demonstrate the program suite's utility for understanding pollutant advection without having to perform statistical analyses first. By combining the data's intensity and temporal and spatial trends (trends that would otherwise be shown in separate box plots, time series, and contour plots, respectively) into one figure (a figure illustrating an event and its before and after events with scales and dimensions), it becomes easier to deduce a pattern (Tufte, 2001; 1998). With a pattern, educated decisions can be made on the direction of subsequent statistical analyses.

To explore the potential link between runoff and fecal indicator bacteria, the LGP was used to display Spearman Rank correlation coefficients (non-parametric measures of correlation that are useful when a jointly normally distributed assumption cannot be made for the variables being correlated; Townend, 2002) between fecal indicator bacteria and salinity (Figure 5). Fecal indicator bacteria concentrations in the water column had a strong negative correlation to salinity during the second Dunes study (as shown by shades of blue in the fourth row, Figure 5). An examination of the raw data collected during the second Dunes study (depicted as animations in the Supporting Information, Figures S1–S3 and as layouts in the Supporting Information, Figures S4–S7) reveals that salinity decreases and fecal indicator bacteria concentration increases during ebb tides. This is consistent with the idea that most of the fecal indicator bacteria comes from sources of freshwater runoff entering Upper Bay, most likely from San Diego Creek and Santa Ana Delhi Channel (see bottom left panel, Figure 1). Figure 5 also shows a weaker negative correlation between fecal indicator bacteria and salinity in several other cases, including sampling sites located near storm drains during the second BYB study, and TC concentrations measured at all sites during the first BYB study and the third Dunes study. The negative correlation between fecal indicator bacteria and salinity at several sites in both the BYB and Dunes field areas is consistent with a runoff source for these organisms.

Conclusions

This study shows that runoff from local storm drains and from creeks draining into Upper Bay is the primary source of fecal indicator bacteria in the water column of Newport Bay. The program suite provides a simple means for visualizing the complex spatial and temporal variability of fecal indicator bacteria in the water column at two marinas (and adjacent areas) in the Bay. Specifically, the concentration of fecal indicator bacteria (and other physical parameters including salinity, pH and turbidity) does not vary much over the sampling grids at a fixed time. However, the concentrations of fecal indicator bacteria exhibit significant temporal variability at multiple time scales—including significant study-to-study variability—with the highest concentrations observed in the studies with the shortest antecedent dry period (e.g., second Dunes study); semi-diurnal and diurnal variability, to a greater extent at the Dunes site and to a lesser extent at the BYB site; and sub-tidal variability, reflected by a slow increase or decrease of fecal indicator bacteria concentrations over the course of our multi-day field experiments.

The results refute the hypothesis that fecal indicator bacteria originate primarily from sources located within the BYB and Dunes marinas. If the marinas were the primary source of contamination in Newport Bay, the concentration of fecal indicator bacteria should be higher in samples collected from the two marinas and lower in the channel sites outside of the marinas, contrary to the trend shown.

The data indicate that runoff from local storm drains and from creeks draining into Upper Bay is the primary source of fecal indicator bacteria in the water column. Once contaminated runoff enters the Bay, fecal indicator bacteria associated with the runoff are transported laterally by the tides (to a greater or lesser extent, depending on location).

The computational tools developed in this study generate user-friendly visualizations that enable rapid interpretation of large multivariate datasets. Furthermore, the computer programs are implemented in a powerful graphics program (Igor, Wavemetrics) and written to be easily adapted to a variety of environmental settings. These rapid visualization programs facilitate first order data interpretation, demonstrate potential for future environmental monitoring studies, and reinforce the utility visualization programs have for analyzing environmental monitoring data (Wood, 1996; Upson, 1989; Edsall, 2000).

Supporting Information Available

The visualization program suite, its source code, animations, and additional figures are available on the Internet at www.urop.uci.edu/framejournal2005supplement.html.

Acknowledgments

Special thanks to Ryan Reeves and Harmony Gates for their mentorship and for coordinating the complex field data collection effort. We would also like to thank the many undergraduates and graduate students who participated in the collection and analysis of the samples. This research was supported by a grant from the University of California Marine Council (UCMarine-32114), the Santa Ana Regional Water Quality Control Board, the City of Newport Beach, and the Henry Samueli School of Engineering Undergraduate Research Fellowship.

Works Cited

California Regional Water Quality Control Board – Santa Ana Region, Fecal coliform Total Maximum Daily Loads for Newport Bay, California, 1999.

Edsall R.M., M. Harrower, and J.L. Mennis. "Tools for visualizing properties of spatial and temporal periodicity in geographic data." Computers and Geosciences 26 (2000): 109–118.

Grunfeld, K. "Integrating spatio-temporal information in environmental monitoring data—a visualization approach applied to moss data." Science of the Total Environment 347 (2005): 1–20.

Jeong, Y., S.B. Grant, S. Ritter, A. Pednekar, L. Candelaria, and C. Winant. "Identifying pollutant sources in tidally mixed systems: Case study of fecal indicator bacteria in Newport Bay, southern California." Environmental Science and Technology 39 (2005): 9083–9093.

Reeves, R.L., S.B. Grant, R.D. Mrse, C.M. Copil Oancea, B.F. Sanders, and A.B. Boehm. "Scaling and management of fecal indicator bacteria in runoff from a coastal urban watershed in southern Calfornia." Environmental Science and Technology 38 (2004): 2637–2648.

Takahashi, H., Y. Hosokawa, K. Furukawa, and H. Yoshimura. "Visualization of an oxygen-deficient bottom water circulation in Osaka Bay, Japan." Estuarine, Coastal and Shelf Science 50 (2000): 81–84.

Townend, J. Practical Statistics for Environmental and Biological Scientists. West Sussex, England: John Wiley and Sons, 2002.

Tufte, Edward R. The Visual Display of Quantitative Information. Cheshire, Connecticut: Graphics Press, 2001.

---. Visual Explanations: Images and Quantities, Evidence and Narrative. Cheshire, Connecticut: Graphics Press, 1998.

Upson, C., T. Faulhaber, D. Kamins, D. Laidlaw, D. Schlegel, J. Vroom, R. Gurwitz, and A. van Dam. "The application visualization system: A computational environment for scientific visualization." IEEE Computer Graphics and Applications 9 (1989): 30–42.

Wood, J., K. Brodlie, and H. Wright. "Visualization over the world wide web and its application to environmental data." Seventh IEEE Visualization 1996 (1996): 81.