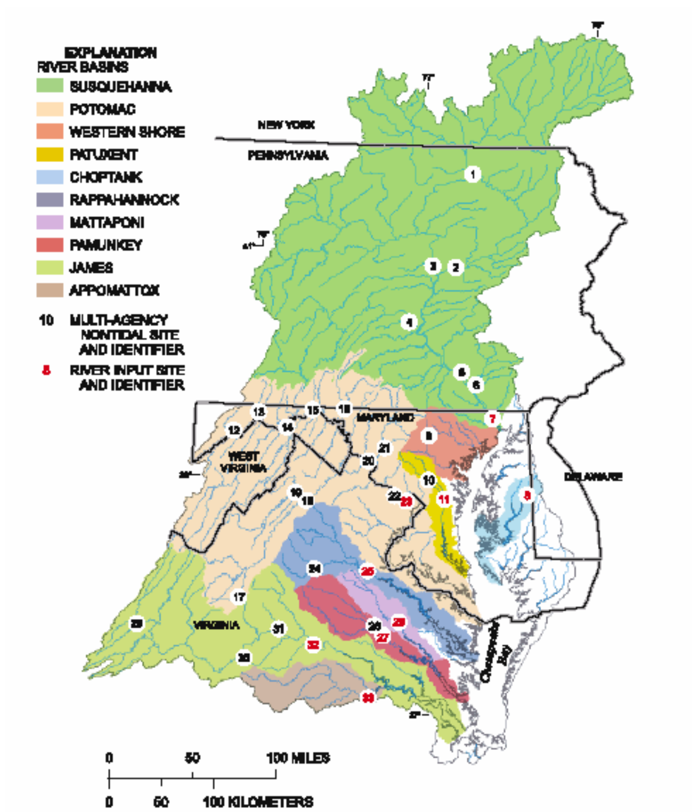
Assessing Progress and Effectiveness through Monitoring Rivers and Streams



June 30, 2005

Report of the Task Force
on Analysis of Non-tidal Water Quality Modeling Results

Scientific and Technical Advisory Committee

STAC Publication 05-005

About the Scientific and Technical Advisory Committee

The Scientific and Technical Advisory Committee (STAC) provides scientific and technical guidance to the Chesapeake Bay Program on measures to restore and protect the Chesapeake Bay. As an advisory committee, STAC reports periodically to the Implementation Committee and annually to the Executive Council. Since it's creation in December 1984, STAC has worked to enhance scientific communication and outreach throughout the Chesapeake Bay watershed and beyond. STAC provides scientific and technical advice in various ways, including (1) technical reports and papers, (2) discussion groups, (3) assistance in organizing merit reviews of CBP programs and projects, (4) technical conferences and workshops, and (5) service by STAC members on CBP subcommittees and workgroups. In addition, STAC has the mechanisms in place that will allow STAC to hold meetings, workshops, and reviews in rapid response to CBP subcommittee and workgroup requests for scientific and technical input. This will allow STAC to provide the CBP subcommittees and workgroups with information and support needed as specific issues arise while working towards meeting the goals outlined in the *Chesapeake 2000* agreement. STAC also acts proactively to bring the most recent scientific information to the Bay Program and its partners. For additional information about STAC, please visit the STAC website at www.chesapeake.org/stac.

Publication Date:
June 2005

Publication Number:
05-005

Cover graphic from Langland et al. (2004).

To receive additional copies of this publication, contact the Chesapeake Research Consortium and request the publication by title and number.

Chesapeake Research Consortium, Inc.
645 Contees Wharf Road
Edgewater, MD 21037
Telephone: 410-798-1283; 301-261-4500
Fax: 410-798-0816
http://www.chesapeake.org

Assessing Progress and Effectiveness through Monitoring Rivers and Streams

Report of the Task Force on Analysis of Nontidal Water Quality Modeling Results

Scientific and Technical Advisory Committee
of the Chesapeake Bay Program

June 30, 2005

Task Force

Donald F. Boesch, Chair
University of Maryland Center for Environmental Science, Chair

Timothy A. Cohn, U.S. Geological Survey

Keith N. Eshleman, University of Maryland Center
for Environmental Science

Thomas J. Grizzard, Jr., Virginia Tech University

James M. Hamlett, Pennsylvania State University

Karen L. Prestegaard, University of Maryland, College Park

Kenneth W. Staver, University of Maryland, College Park

Donald E. Weller, Smithsonian Environmental Research Center

STAC Publication 05-005

Contents

Executive Summary

Water quality monitoring of rivers draining to tidal waters of the Chesapeake Bay is a critical component of the restoration of the estuarine ecosystem and its watershed. Reducing the nutrient and sediment loads to the Bay that are largely responsible for degrading the health of the estuarine ecosystem, is the keystone goal of the Chesapeake Bay Program (CBP). Meeting the environmental quality objectives of the Chesapeake 2000 Agreement will require reductions in nitrogen and phosphorus of 48 and 53%, respectively, from loads estimated for the mid-1980s. Because approximately three-fourths of the nutrients and virtually all of the sediments are from diffuse rather than point sources, measuring changes in the loads emanating from the watershed is essential to gauging progress in meeting load reduction targets and determining the effectiveness of management practices.

Although nontidal water quality monitoring has been conducted for 20 years or more at fixed sites throughout the watershed, including nine River Input Monitoring (RIM) sites located above the tidal influence in the major tributary rivers, interpreting the trends in nutrient and sediment loads is not simple. Annual freshwater discharges into the Chesapeake Bay have varied by a factor of three and flows were unusually variable during the past 15 years when management practices were being applied by CBP participants. Variable discharges greatly affect loads, making it difficult to assess trends. Tests for trends in flow-adjusted concentrations of nutrients and sediments show that significant reductions have occurred in many, but not all, of the major rivers draining to the Bay. However, there is confusion among scientists, managers, the press and the public as to the magnitude of these reductions and how they compare to model projections of progress toward the Chesapeake 2000 nutrient and sediment reduction goals.

The Scientific and Technical Advisory Committee (STAC) assembled this Task Force to advise the CBP on interpreting nontidal water quality monitoring results with regard to management decisions, more effective methods of data analysis, and the appropriate integration of monitoring results and modeling activities. Although, the Task Force's assessment has many specific findings and recommendations, the following main conclusions are drawn:

1. Nontidal water quality monitoring is an indispensable technical tool for guiding the Chesapeake Bay Program. Monitoring serves multiple purposes, including: estimating nutrients and sediment loads; quantifying trends in concentrations that potentially reflect effectiveness of control practices; calibrating the CBP Watershed Model and verifying and improving its performance; and revealing the effects of climatic variability on loading. Greater use and interpretation of nontidal water quality monitoring results should be explored beyond estimating loads and trends, particularly more effective integration with modeling, evaluation of management effectiveness, and assessment of climatic and landscape factors.

2. Evaluating progress in achieving nutrient and sediment reduction goals should be based to the greatest extent possible on direct observation through watershed and point source monitoring rather than on model estimates. Continued analytical refinements are desirable, but the present method of quantifying trends in flow-adjusted concentrations, which corrects for the effects of flow variability, provides the best currently available indicator of long-term responses to management activities.

3. The spatial density of nontidal water quality monitoring is inadequate for determining the effectiveness of control actions being taken within the watershed, particularly when considering the great costs and uncertain effectiveness of implementing these actions. The monitoring network should be strategically enhanced to include watersheds that can deliver high loads to the Bay, especially urban and agricultural regions within the Coastal Plain, and networks of sites within major river basins. Long-term monitoring results from other sites and studies within the watershed should be used to enhance understanding of load dynamics and the effects of land-use practices on loads.

4. Methods of assessing water quality and loading trends should be further developed, compared and contrasted to refine our ability to estimate meaningful changes in concentrations and loads, particularly with regard to climate-related variability, trends within basins, smaller watersheds, longer time periods, and sensitivity to sampling frequency.

5. Monitoring results and model estimates should be more clearly and accurately communicated to managers and the public. Such communication should always make clear the degree to which conclusions regarding progress or effectiveness in achieving restoration goals are based on actual observations or modeled projections. Recent efforts with the CBP to differentiate quantification of the "state of the bay" from the "state of the bay restoration" are consistent with this advice. Moreover, to the degree possible, they should identify the uncertainties involved in results and estimates.

6. Periodic independent review of the design, methods, analyses and interpretations of the nontidal water-quality monitoring program is important for ensuring reliable results that are well accepted in the scientific community and useful to managers. Such reviews should be conducted by competent scientific peers and information users, particularly prior to the release of major interpretive reports.

7. Although the distinction between monitoring results and modeling estimates should be recognized and clearly maintained, watershed modeling and monitoring activities should be better integrated, along with research results, within an adaptive management framework to: reveal key uncertainties; provide focus for strategic research; improve model accuracy and interpretation of the observations; and contribute to evaluation of the effectiveness of nutrient and sediment control efforts.

Introduction

"We are drowning in information, while starving for wisdom. The world henceforth will be run by synthesizers, people able to put together the right information, at the right time, think critically about it and make important choices wisely." E.O. Wilson

Background to Task Force Study

Reducing the loads of nutrients and sediments flowing and discharged into the estuary is the keystone goal of the Chesapeake Bay Program (CBP). The Chesapeake 2000 Agreement (Chesapeake Executive Council, 1999) committed its signatories to "correct the nutrient- and sediment-related problems in the Chesapeake Bay and its tidal tributaries sufficiently to remove the Bay and the tidal portions of its tributaries from the list of impaired waters under the Clean Water Act by 2010." Using computer models of the estuary and the Bay's watershed, it was determined that annual average loads of total nitrogen (TN) must be reduced to 175 million pounds and annual average loads of total phosphorus (TP) reduced to 12.8 million pounds to achieve ambient water quality criteria for dissolved oxygen, water clarity and chlorophyll levels in the Bay and its tidal tributaries (Koroncai et al., 2003). These target loads are approximately 48 and 53%, respectively, lower than estimated for 1985, the baseline year for the restoration effort.

Diffuse, or nonpoint, sources dominate the loads of nutrients to the Bay, comprising 74% of the TN in 1985 and 79% in 2002 and 66% of the TP in 1985 and 78% in 2002, based on estimates by the CBP. The diffuse sources emanate primarily from agricultural activities, urban runoff, atmospheric deposition onto tidal waters and the watershed, and erosion. The fluxes from these diffuse sources are not only notoriously difficult to control, but they are also far more difficult to measure than point source discharges. The CBP uses models to determine strategically the abatement practices needed to achieve restoration objectives and relies on a monitoring program to gauge progress in Bay restoration. A particularly important component of the monitoring involves measuring the discharge of nutrients and sediments by nine major river systems that drain to the Bay, but some smaller rivers are also monitored.

Fueled by concerns that improvements in water quality in the estuary are not yet apparent despite more that 15 years of concerted effort to reduce nutrient loads, criticisms have been leveled at the CBP through books (Ernst, 2003), the news media, and advocacy organizations. These criticisms contend that the CBP assessments of progress in reducing nutrient loads have been overly optimistic and misrepresentations of actual conditions, in that they were based on model estimates rather than documented outcomes. The lack of clear evidence of progress based on watershed monitoring results was also raised in high-profile press coverage (Whoriskey, 2004). As a result of confusion about the actual progress that has been made, an investigation by the Government Accountability Office was launched in 2004 at the request of Congress.

These developments have inspired a broad debate and a reappraisal within the CBP on how monitoring and modeling are used to assess progress and effectiveness in reducing nutrient loads to the Bay. Although these two activities are quite different, modern management requires the comparison of monitoring results and model predictions to realistically track progress and appropriately inform management regarding the effectiveness of actions through a process called adaptive management (e.g. see U.S. Commission on Ocean Policy, 2004).

The Scientific and Technical Advisory Committee (STAC) undertook this assessment of the analyses and interpretation of non-tidal water quality monitoring results in order to enhance both their scientific rigor and utility in management decisions. Actually, STAC has long had an interest in the integration of monitoring and modeling in the Bay and its watershed, previously producing two reports on the subject (STAC, 1997, 1998). The assessment reported herein addresses the appropriate uses and integration of non-tidal monitoring and watershed modeling, but will not provide in-depth evaluation of the present watershed model or the basinwide nontidal monitoring network, which are the subjects of separate STAC-sponsored reviews (Band, et al., 2005; STAC 2005).

Three Major Issues Considered by Task Force

The Task Force attempted to address a series of questions grouped under three interrelated issues:

1. **Alignment of monitoring and data analyses with management decisions**. What are the key management decisions (both perceived and potential) that should be informed by monitoring results? How should data analysis be structured to provide the information, including indicators of progress and effectiveness, needed to support adaptive management decisions? How can uncertainty and confidence limits best be expressed to decision makers to provide appropriate understanding of the limitations associated with the monitoring program and the analyses and interpretation of results?

2. **Effective methods of data analysis**. What are the most appropriate, robust and effective methods of data analyses to address the management questions? How can the methods be made transparent and interpretations and uncertainties be made clear? How should external experts and reviewers be engaged in the process of data analysis and interpretation?

3. **Integration of monitoring and modeling**. How can monitoring results be used to improve the model performance? How can models be used to improve the interpretation of monitoring results, e.g. through comparisons of observations and predictions? What are the best approaches for integration of monitoring and modeling in assessments of progress toward achieving goals and effectiveness of management actions? What institutional adjustments are required to accomplish more effective integration and use of these technical tools in decision-making processes?

Approach to the Study and Report

The objective, tasks and approach were discussed with appropriate technical managers in the U.S. Environmental Protection Agency and U.S. Geological Survey and modified appropriately. A Task Force was assembled to include both members of STAC and outside experts. The Task Force includes individuals with experience in statistical analysis, hydrology, geochemistry, nonpoint source technologies, and science synthesis.

On April 15, 2005, the Task Force engaged in a group consultation with managers and technical specialists within the USGS and EPA to gain an understanding of the current practices, ongoing improvements, and perceived needs. This meeting was structured to maximize the open exchange of information and ideas. We greatly appreciate the information provided by, and full cooperation of, Scott Phillips, Steve Preston, and Jeff Raffensperger of the USGS and Gary Shenk of the EPA Chesapeake Bay Program Office.

The Task Force met again on May 6 for further review and discussion among its members on the draft reports of each main task, to prepare a listing of key findings/recommendations, and to continue work on a draft of this report. Written contributions by Task Force members were integrated and sequentially revised to produce this final report.

Agency representatives provided comments on a draft of this report to ensure factual accuracy and clarity, but the findings and recommendations of the report are solely those of the Task Force. We thank STAC members James Lynch and Joseph Bachman for providing helpful review of the same draft.

Alignment of Monitoring and Data Analyses with Management Decisions

Informed Management

To be done successfully, the complex tasks of water resources and ecosystem management are usually dependent on the actions of three groups of people: a group that makes regular observations in the environment (i.e., *monitoring*), a group that makes forecasts based on conceptual understanding of the system (i.e., *modeling*), and a group that actually makes decisions and takes actions (i.e., *management*). Of course, hypothesis-driven *research* is also essential, helping to elucidate the factors that control the underlying phenomena. Water quality monitoring and modeling have historically played very significant roles in traditional water pollution abatement, as well as in contemporary management of freshwater and estuarine ecosystems. There are many examples (e.g., Lake Washington, Lake Mendota, etc.) that illustrate the important role played by water quality monitoring and modeling in management of freshwater eutrophication.

Traditionally, water quality monitoring, modeling, and management efforts were largely conducted in isolation; models were developed largely on the basis of the availability of data, whereas management decisions were made based on the resultant model forecasts. The more recent mantra in water resources management is *integration* such that modeling and monitoring activities should work closely together, rather than in isolation. Management decisions and actions can thus be based on overall system understanding and on forecasts from models that are developed, tested, and calibrated using actual field data from a carefully-designed monitoring network that is tailored to the needs of the specific model (Duckstein et al., 1985). Carried a step further, this integration leads to *adaptive management*, in which the actions, and even goals, are adjusted based on the interpretation of carefully monitored outcomes of previous management actions (NRC, 2004).

Integration of monitoring and modeling in water quality management has both practical benefits and limitations. One important benefit is often a reduction in cost associated with both tasks, since model development and monitoring results are somewhat "optimized" when treated as complementary components of a larger task: simpler models that require fewer (i.e., cheaper) monitoring data are preferred over more complex models that have great (i.e., expensive) data needs. Similarly, collection of data that are not relevant to a particular model may be avoided, thus saving limited monitoring resources. Data serve multiple purposes, of course, and there is a cost associated with optimizing the sampling to address the needs of modeling; the broader objectives of monitoring must also be considered.

The integration among monitoring, modeling, and management is implicit in the Chesapeake Bay Program (CBP, 1999), although the means and effectiveness of the integration are subject to question. Monitoring activities within the CBP address tidal and nontidal water quality, living resources, and land use/land cover issues; but only nontidal water quality monitoring is considered here. The CBP employs several management-oriented modeling approaches (Linker et al., 2002). These include a Watershed Model (WM) and a Chesapeake Bay Estuary Model (CBEM), which are linked—the WM delivering nutrient and sediment loads to the CBEM. While models used in the CBP can produce results representing past and current conditions, their particular value is in forecasting future conditions based on different scenarios of changes in land use and management practices. Both monitoring data and model results can generate similar products (for example, annual nitrogen loads), therefore, it is important to consider what the monitoring data and model results each represent, how they were derived, and how they can each be used. These issues are addressed later in the section entitled "Integration of Monitoring and Modeling."

The purpose of this section of the report is to review the current alignment of nontidal water quality monitoring and analyses of its results data with management decisions in the Chesapeake Bay Program (CBP). In this regard it is important to remember that monitoring results may have many different uses.

Results from environmental monitoring can be applied to a variety of objectives (Boesch, 2000)

Trends	Determining the present status and past trends in an environment or resource
Goals	Gauging the degree to which specific goals have been met
Resources	Assessing the supply of resources for management of their uses
Precaution	Assuring precautionary protection by determining whether some action level is exceeded
Models	Applying to the calibration and verification of ecosystem, water quality, resource management, or exposure models
Research	Providing the context of hypotheses tested by scientific research and the extrapolation of research results
Public information	Informing the public about the world in which we live in order both to account for government programs and to promote citizen stewardship
Emergencies	Supplying background information for response to unforeseen effects and emergencies

One important focus of nontidal water quality monitoring is the determination of trends in the delivery of water, nutrients and sediments from the watershed to the tidal waters of the Chesapeake Bay. Another should be the degree to which goals for reductions in the delivery of nutrients and sediments, and thereby the effectiveness of management actions, are being met. In this section, we address the important aspects of water quality management of Chesapeake Bay that can, and should, be informed primarily by monitoring data, independent of any model or forecasting tool.

The Nontidal Water Quality Monitoring Network

The current status and past trends of the amount and quality of water delivered to the Chesapeake Bay have been monitored at 33 sites in the nontidal portion of the Chesapeake Bay basin and are analyzed and interpreted each year and periodically reported by the U.S. Geological Survey (Langland et al., 1999, 2004). These 33 sites represent a sparse network for such a large watershed (Figure 1). They include nine critical River Input Monitoring (RIM) sites located above the head of tide in larger tributaries and a rather small number of additional sites in the Susquehanna, Potomac, and James catchments. Nontidal streamflow and water quality for most of the Coastal Plain (estimated via model predictions to contribute 38% and 40% of the TN and TP loads, respectively) that includes urbanized portions of the Baltimore-Washington-Richmond corridor and Tidewater Virginia and almost all of the agriculturally intensive Eastern Shore are not included in this monitoring program. Estimates derived from the WM suggests that 50-60% of the nutrient reduction goals will have to be met within this presently unmonitored area, much of that from nonpoint sources (Gary Shenk, personal communication). At present, these nonpoint source loads have to be estimated from models.

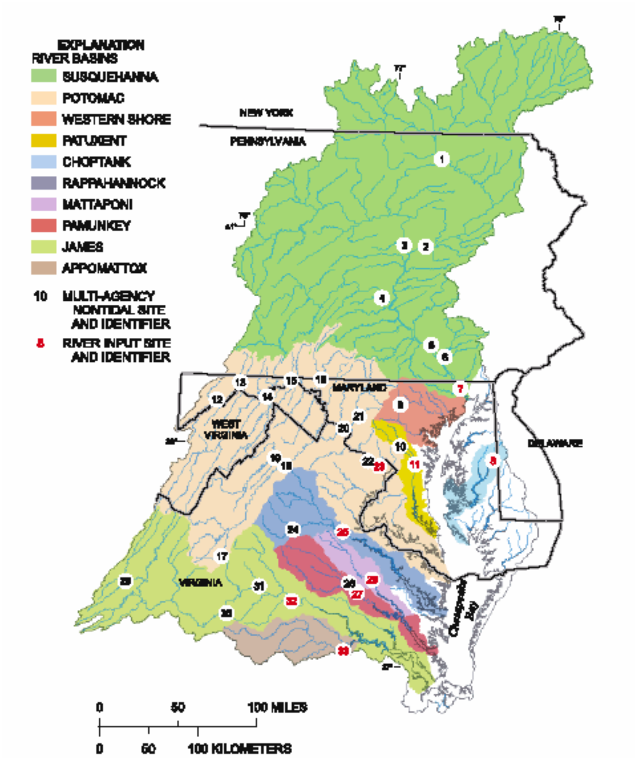


Figure 1. Location of the 33 sites used for analyses of nontidal water quality trends; red numbers represent RIM sites (Langland et al. 2004).

Most of the monitoring sites are USGS gauging stations, where gauge height is continuously monitored with devices that can accurately record flow depth or water elevation. These data are converted to stream discharge values by use of “rating curves,” which are statistically robust relationships between gauge height and measured discharge at each site. Water quality samples are collected within the vicinity of the stream gauging station according to a sampling scheme designed for the purpose. In the case of sediment and nutrient loads to the Bay, samples are collected at both high and low flows to evaluate nutrient and sediment loads at a range of flows. These streamflow and water quality data are then analyzed using various statistical techniques to evaluate annual loads or other water quality parameters. The streamflow and water quality data are available online and are analyzed in reports periodically produced by the USGS, the most recent at the end of 2004 (Langland et al., 2004). While care is taken to ensure that the processes for actual measurements are consistent over time, interpretations of these data do change as a result of trends that might emerge or disappear as the record lengthens or as improvements in data analysis and statistical procedures evolve.

It might appear to be a relatively simple procedure to present and interpret the monitored streamflow and water quality data. Although data collection and analysis procedures are relatively routine, interpretation of the results involves an understanding of the hydrological and nutrient flux systems, which requires data collected over many years. Trends in water quality would be easy to interpret if hydrological conditions remained constant from year to year and only the sources of nutrients and sediments changed. The Chesapeake Bay watershed, however, demonstrates significant hydrological variability from year to year. An unusually wet year, with twice the annual precipitation, can follow an unusually dry year (Figure 2). In fact, the 15 years since 1990, during which almost all of the concerted efforts to reduce nutrient loading took place, have been a period of unusually variable streamflow into the Chesapeake Bay. Unless there is drought for the remainder of this very wet year, 2005 will be one of seven wet years during this period with annual discharge falling in the top quartile of the long-term record and only three years during this 15-year period have had river discharges falling into the “normal” central two quartiles.

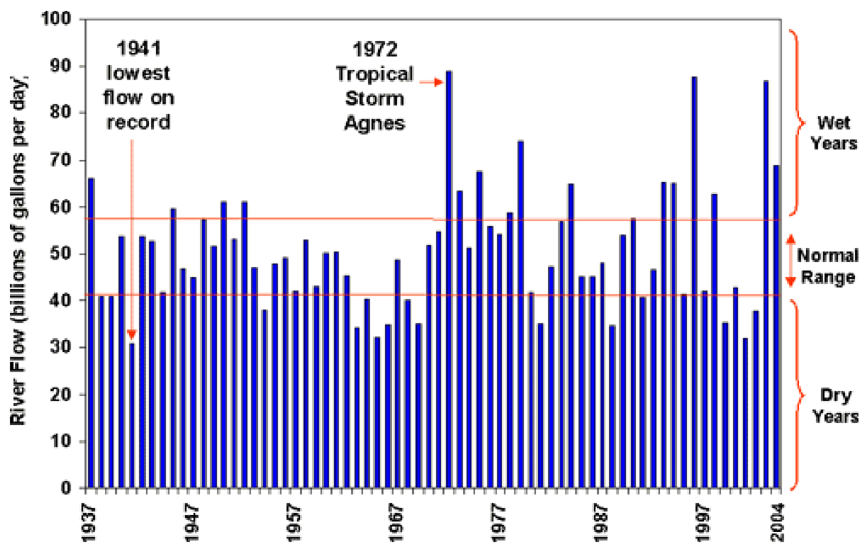


Figure 2. Annual variations in mean flow entering the Chesapeake Bay. The “normal range” is defined as including the years in the long-term record falling between the 25th and 75th percentile.

This hydrological variability greatly influences both the concentrations and loads of nutrients and sediments in the streams; large storms generate high loads and small loads are carried during periods of low streamflow. However, the variable flows also affect the concentrations of nutrients and sediments through complex processes of mobilization and dilution. Therefore, the USGS assesses long-term trends by computing flow-adjusted concentrations (FAC) based on estimated daily loads from a 7-parameter, log-linear regression model (ESTIMATOR, Cohn et al., 1992). Computing FAC is a way of factoring the hydrological variability out of the long-term trends. Although the data actually collected are for flow and concentration, water-quality results, computed by the methods reported in Langland et al. (2004), are presented as observed, flow-weighted (FWC, approximating the annual concentration), and flow-adjusted (FAC) concentrations and loads. In this most recent USGS report, only FAC is subjected to statistical testing of trends—both raw concentration data and FWC estimates are difficult to interpret because the sampling of concentration is biased toward higher flows (see page 15 of this report).

The complexity of the flow-concentration-load relationships leads to confusion among managers, the press, the public and even scientists about the meaning of these monitoring results from nontidal water quality data analyses. For example, Langland et al. (2004; Fig. 11) report declining trends in FAC of nitrogen, phosphorus and sediments, many statistically significant, for the larger rivers (Susquehanna, Potomac, and James, except for P in the Potomac); however, because of the high interannual variability in flow there is no apparent trend in nutrient loads, which were at near-record levels in 2003 (Langland et al., 2004, Figure 12). This underlies the disconnect between trends appropriately discerned from concentrations and conditions actually observed in the estuarine ecosystem, which responds to nutrient loadings. Citizens and even managers are left puzzled, then, about why we are not seeing more improvements in the Bay if we are making such progress with source reductions.

The lack of clarity and transparency was also apparent in news media criticisms of inappropriate representation of model results as an indication of progress. The lack of a trend for decreasing phosphorus concentrations was used to call into question the interpretation of monitoring results, but was based on observed, not flow-adjusted concentrations (Whoriskey, 2004). At the time of the media attention to this issue, USGS scientists were evaluating the most appropriate techniques to assess changes in water quality and had not yet completed an assessment report on results later than 1998 (Langland et al. 1999). They had not reached consensus on the magnitude of changes observed that could be explained to the press at that time.

Scientists may also question the reliance on statistically computed FACs without a full understanding of the complex, mechanistic interrelationships among sources, flows and concentrations. Regional scientists most experienced in research on the processes involved raise concerns about whether important seasonal dynamics and potential underestimation of loads during extreme events are adequately taken into account.

Key Requirements

Because improvement of water quality conditions is often a primary goal in ecosystem management, water quality data obtained through monitoring often represent an early, easily interpreted, and useful indicator of overall ecosystem health in their own right. It is worth reviewing the variety of key management decisions that can and should be informed either exclusively or principally through analysis of monitoring results:

- characterizing waters and identifying trends or changes in water quality with time;
- identifying existing or emerging water quality problems;

- gathering information for design of pollution prevention, abatement, or restoration programs;
- determining if the goals of specific programs (i.e., pollution prevention, abatement, or restoration strategies) are being achieved;
- responding to emergencies (e.g., spills, floods, etc.); and
- providing information to agencies and the public for use in decision-making about resource use (e.g., beach closings, limits on recreational use, fishing bans, etc.).

As described by Ernst (2003) in his recent book *Chesapeake Bay Blues*, a variety of ecological and political complexities continue to represent significant challenges in understanding the health of the ecosystem, in promulgating regulations that could help achieve restoration goals, and in implementing management strategies fairly and consistently across the entire ecosystem represented by Chesapeake Bay and its watershed. The Chesapeake Bay Program (CBP) sets collaborative goals and the U.S. Environmental Protection Agency has primary jurisdiction for only a few of the management decisions and actions noted above; most of the other decisions and actions are essentially being left to the states or to individual localities to implement. The CBP does not even have direct responsibility for water quality monitoring of non-tidal tributaries, forcing it to rely on state, or state/federal, cooperative monitoring programs to provide the information that it needs to achieve several of its primary goals (Chesapeake Watershed Partners, 2004). This is a significant issue, since any environmental restoration program is ultimately dependent upon the monitoring results to document water quality improvements; the present reliance on others to provide monitoring results has some advantages and disadvantages (discussed below). Regardless of the source of the monitoring data, their use could be greatly improved to facilitate management decisions being made both inside and outside of CBP. Recommendations for improving these analyses are as follows:

1. Maintain the primacy of monitoring data and analyses. While monitoring and modeling must be integrated for certain purposes (i.e., model calibration, verification, etc.), it is essential that the collection and statistical analysis of monitoring data (specifically the water quality data) be maintained in ways that are “beyond reproach” and “uncorrupted” by modeling results. There are several reasons for this. While it is fairly well accepted that monitoring networks and statistical analyses must be flexible (i.e., as stations in the network are added or deleted, as additional years of data are collected, as analytical/statistical methods are changed, etc.), the primacy of the monitoring data as the “gold standard” should never be needlessly compromised for reasons of administrative expediency. Since the monitoring data are usually the best (and, in some cases, the only) reflection of the efficacy of the overall management effort, the value of these data increase dramatically as time passes. Even if a specific management action is later deemed to have been unsuccessful, the value of the monitoring data can be maintained as a legacy to provide the foundation for future management decisions.

2. Maintain the distinction and independence of monitoring results and analyses from modeling results and interpretations. The goal of maintaining the primacy of the monitoring data might be further advanced by the creation of a "bright line" of distinction between the analysis and interpretation the monitoring data (the "facts") and the production and interpretation of modeling results. This should be similar to the way in which the judicial system separates evidence (data) from legal arguments. The advantage of such a bright line is that the independence of the monitoring data is easier to maintain, thus avoiding questions about bias, overstatement of progress, etc. Having said this, it is also recognized that more effective integration of monitoring data and modeling results is required in the CBP for assistance in interpretation of the monitoring data, model refinement, and adaptive management (Boesch, 2004). Furthermore, some estimates of progress and effectiveness will have to rely on a combination of monitoring observations and model estimates (e.g. to address loads from the unmonitored portion of the watershed). However, the degree to which conclusions are drawn based on direct observation or model estimates should always be made crystal clear. In earlier State of the Bay reports (CBP, 2002) estimates based on models combined with observations for point-source measurements were represented in a way that conveyed a false certainty in progress in nutrient and sediment load reductions. The most recent report (CBP, 2004) corrects this by clearly indicating which charts are based on model estimates. The move within the CBP to differentiate quantification of the "state of the bay" from the "state of the bay restoration" is also consistent with our advice.

3. Perform standardized, statistical analyses of monitoring data and release results to the public on a periodic basis. It must be recognized that graphs or other displays of monitoring results can have tremendous impacts on managers and the public alike. Perhaps the best example is the simple, yet influential, "Bernie Fowler Sneaker Index" that has been used to convey changes in water transparency in Chesapeake Bay during the last 50 years. Given these possible impacts of monitoring data, there is an obvious temptation to release information quickly to either drum up support or highlight ostensible progress toward a particular goal. Achieving the ambitious CBP nutrient reduction goals is likely to be very gradual (occurring over decadal, not annual, time scales) and occasionally non-monotonic, owing to a number of factors. These include (a) large groundwater residence times and other factors that delay the impact of management actions on the delivery of nutrients and sediments to the Bay; (b) climate variability, especially years with above average runoff; (c) rapid population growth and associated land use changes in the watershed; and (d) delays in actually implementing management actions. Given the desire to sustain progress (and support) over a long period of time, it is important to educate the public and decision-makers in this regard. In the same vein, the plethora of progress reports released in the last five years, that emphasized process rather than long-term outcomes, has probably conditioned the receivers of this information (through the media) to think that progress should be occurring more quickly than it has. Restoration of Chesapeake Bay and other ecosystems cannot be effectively evaluated at the same frequency as, for example, a stock portfolio. For that reason, status and trend reports on nontidal water quality should be undertaken on an appropriate multi-year basis, sufficiently in advance of management reassessments (e.g. the planned 2007 reassessment of Chesapeake 2000 Agreement). While recognizing that annual load estimates must be regularly provided to program managers, key milestone reports, such as the USGS nontidal water quality trend reports, should be externally peer-reviewed for statistical analysis and interpretation prior to their publication. Original monitoring data should, of course, remain accessible to the scientific community, managers and the public on a near real-time basis.

Effective Methods of Data Analysis

The nontidal water-quality monitoring network plays a critical role in planning and accountability for the Chesapeake Bay Program. Much has been asked of the agencies and individuals charged with network operation and interpretation of results. Over time, the network has been used for a variety of tasks, including:

- development of estimates of constituent loads delivered to the Chesapeake Bay;
- provision of data for calibration and verification of the Watershed Model (WM);
- provision of regular assessments of water quality in the major tributaries; and
- provision of long-term time series data to support assessments of water quality trends, and therefore, restoration progress as measured by watershed constituent load reductions.

Unfortunately, the monitoring network has been largely developed and optimized over time to achieve only the first of the principal goals, namely the development of annual loading estimates from the major tributaries to the Bay. A major infusion of resources will be necessary to ensure that the nontidal monitoring network can achieve all the results that scientists, managers, and decision-makers expect. With the scale of the cost of achieving the Chesapeake 2000 nutrient reductions exceeding $20 billion (Chesapeake Bay Watershed Blue Ribbon Finance Panel, 2004), it is worth investing significant resources for monitoring so as to better determine if management and program approaches are effective in achieving the keystone goal of the Chesapeake Bay restoration.

Current Trend Analyses

It is in the best interest of the CBP to continue to seek the most effective tools to support the analysis of nutrient and sediment trends in rivers and streams draining into the Bay. Trend analyses, particularly if ways can be found to successfully normalize for flow condition, may be used to reveal long-term progress in restoration.

It seems useful to continue reporting observed concentrations, flow-weighted concentrations (FWC) and flow-adjusted concentrations (FAC) to provide comparability with prior analysis and reports. Because of the statistical limitations discussed by Langland et al. (2004) (see page 15 of this report), the use of FAC seems more appropriate for statistical testing of trends; however, FWC is an informative statistic, commonly used in other long-term studies and probably the best estimate of average concentration. Patterns in FWC and loads should be presented graphically; these results reflect different phenomena that are important for management interpretations. However, in the most recent report, load estimates are represented only for the aggregate of RIM sites. Presentation of time series for annual load estimates separately for each of the RIM sites would be more useful to those working to implement tributary strategies for nutrient load reductions and for the interpretation of monitoring results from tidal waters in the upper Bay and tributaries. It is also necessary to ensure that additional effort is directed towards explanations of uncertainty and apparent changes in trends associated with hydrometeorologic phenomena. Information conveyed to non-scientific, non-technical audiences is particularly susceptible to misunderstanding and misinterpretation, particularly where the underlying trends are difficult to discern because of the scale of natural variability.

Langland et al. (2004) identified some concerns about whether the current sampling and data analysis protocols are appropriate for satisfactory determinations of trends in either observed constituent concentrations or flow-weighted constituent concentrations in the major tributary rivers. They cite a bias toward storm sampling (because the sampling strategy is optimized to support ESTIMATOR load estimates instead of concentration trends) as a reason for not continuing with analysis of trends in observed data. This suggests that it may be instructive to analyze baseflow concentrations separately as a complement to FAC trend analysis. Given that observed data are generally easiest to explain to the public and decision-makers, perhaps the sampling strategies should be revisited. It will take time (and money) to produce satisfactory datasets, but that should not be the only reason for abandoning this approach. Larger datasets, which are distributed across a range of flow conditions, already exist at Chain Bridge in the Potomac, and might be useful in evaluating sampling bias.

The 2004 report also cites concerns about conducting FWC trend analyses based on the quotient of estimated monthly load (from ESTIMATOR) and observed monthly flow. As with concentration data, there may be more information available at Chain Bridge to evaluate the utility of a FWC trend analysis based on a flow-compositing sampling strategy for storms and a regular time series strategy for non-storm conditions.

Exploration of New Trend Analyses

Several observations may be made regarding the analysis of existing datasets. First, it seems desirable to use longer time series data where they are available. Where data of sufficient quality exist, analyses that deal with longer time periods will help to put the results from the relatively short 1985-2004 period in context. Furthermore, such data would provide insight on the degree to which time-series data for streamflow exhibit stationary behavior (see below).

Additional work needs to be done to develop the best flow-adjusted concentration (FAC) trend estimate from the ESTIMATOR regression model. We suggest several areas to explore for potential improvements:

- Further evaluation should be performed of the inclusion and interpretation of the quadratic term from the ESTIMATOR model to estimate trends in FAC.
- In some cases it might be appropriate to incorporate step changes into the time series analysis that would better account for single management actions (e.g., detergent phosphate bans, construction of sewage treatment plants, etc.) and additional polynomial curve fitting.

- Attempts should be made to correct for sampling bias in order to estimate concentration trends independent of the ESTIMATOR model.
- New data analytical approaches proposed by Greg Schwarz of the USGS appear to be promising, but require more detail in order to be evaluated.
- Further exploration of Autoregressive Integrated Moving Average (ARIMA) and other statistical models should be undertaken to assess their applicability to interpretation of water quality trends.
- ESTIMATOR assumes a set of relationships between load and discharge over a nine-year time window. However, land uses and land cover or the trapping efficiency of dams may change over such a period. Furthermore, while it appears that some of the time series are stationary, many of the runoff datasets from Chesapeake Bay watershed sites exhibit significant non-stationarity, associated mostly with climatic variations. The effects of non-stationary behavior on the trend analyses of water quality should be examined and quantified.

Other Nontidal Monitoring Data Analyses

There is a need to conduct broader analyses of the nontidal monitoring data and the associated hydrometeorologic variables. Trend analysis is valuable and plays a central role for the CBP, but there is a need to assess temporal variation more generally, recognizing that trends on the decadal scale are only one component of temporal variation. In particular, a more complete analysis and presentation of the range of variability due to climatic conditions would be useful. This would help to place the observations of a high or low nutrient flux seasons or years in a hydrometerological context and would forestall over-interpretation of data from single years.

Efforts should continue to push the analysis of the existing data as far as possible, in particular with regard to quantifying loads, trends, and other indicators of progress. Data analysis should proceed independently of the WM development, but analysis results should eventually be compared with the model projections (while always clearly differentiating observational results from model estimates as discussed previously).

Periodic comparisons of model results to monitoring data should be conducted. In the course of the committee deliberations, it was noted that the WM calibration process has always been performed independent of the RIM loads computed with ESTIMATOR. In addition to comparing model results and monitoring data, it would be useful to compare ESTIMATOR loads at the RIM stations at an appropriate frequency (monthly, seasonally, annually) to the WM output. This would be a useful check on the variability of model predictions.

Due to the relatively complete, continuous gauging and capture of aggregate loads above the fall line, data from nine RIM stations are primarily used as input to the ESTIMATOR method that is subsequently used to estimate annual loadings and compute flow-weighted concentrations (FWC) and flow adjusted concentrations (FAC) of water quality constituents. While trend analyses of monitoring data are appropriately focused on RIM stations and the CB watershed as a whole, greater analysis of other "within-basin" stations should be conducted to support interpretations of data from the RIM sites. An analysis of the regional representativeness and biases associated with use of the present 33 monitoring station network should be performed, since data from stations that are close together are expected to be highly correlated. Any changes to the monitoring network should be made only after taking the regional representativeness of the sample of stations into account.

Better interpretation of time series data and trends of water quality from individual watersheds with respect to past management actions would be desirable. There has been too much reliance on "global" trend analyses, often done with little supporting documentation, at the expense of more carefully crafted trend analyses for individual non-tidal systems (e.g., Patuxent River, Potomac River, Susquehanna River, etc.).

Approaches should also be sought to exploit shorter (but already existing) data records from monitoring sites not currently used in the nontidal monitoring network. Attempts should be made to construct an inventory of such small catchment and stream data records in the Chesapeake Bay watershed. Small catchment or small watershed studies may already exist that have attempted to relate loads to watershed activity at a scale where effects may be reasonably detected. Analysis of such datasets may be useful in explaining historical anomalies in the RIM station record.

Locations already exist in the Chesapeake Bay watershed where composite sampling strategies have been applied to the development of annual load estimates from a variety of small, medium and large watersheds. Anecdotal evidence indicates that, at least at the large watershed scale, load computations from composite sampling programs and computations from ESTIMATOR yield statistically indistinguishable results for annual loads. Efforts should be made to establish collaborations with other investigators so that such information can be used to explore the limits of ESTIMATOR at time scales of less than a year (monthly? seasonally?). An example of where this might be done most successfully at the large watershed scale is the Potomac River at Chain Bridge (which is also a RIM station). The river cross-section is sufficiently well mixed (both laterally and vertically) at Chain Bridge to allow the use of an automated, point-intake sampler. A database that exists for the 1982-2004 period would be useful for comparisons of monthly, seasonal, and annual load estimates.

With regard to smaller watersheds, there are a number of historical or on-going programs that have been designed to assess the impacts of watershed activity on constituent export. A number of investigators in the watershed have developed long records of hydrology and chemistry for basin sizes ranging from a few acres to 500 sq. mi. The Smithsonian Environmental Research Center, Virginia Tech Occoquan Laboratory, University of Maryland Center for Environmental Science, and the Leading Ridge Experimental Watersheds in Pennsylvania are reasonable places to begin such an inventory.

Expansion of Nontidal Monitoring Network

As was noted earlier, a principal purpose of nontidal water quality monitoring has been to support the development of load estimates from the free-flowing tributaries to the Bay. The RIM network has performed satisfactorily with regard to estimating annual loads from large watersheds. The other requirements being placed on nontidal monitoring will require changes in monitoring strategy as well as greater temporal and spatial coverage. Fortunately, state and federal agencies and river basin commissions recently signed a Memorandum of Understanding to expand the Non-tidal Water Quality Monitoring Network (NTWQMN) by as many as 100 sites selected from a list of 188 candidate stations that included 115 existing stations and 73 new locations, identified as characterizing tributary strategy basin segments, areas with the highest nutrient delivery or representative of the overall range of conditions in the watershed. Availability of fiscal resources will undoubtedly limit this expansion, which has been review by another panel of the Scientific and Technical Advisory Committee (STAC, 2005). Nonetheless, from the perspective of this Task Force, the following suggestions are offered.

- ***Currently un-monitored areas.*** The existing program has been deficient (for reasons of methodology and resources) in obtaining loading data from currently unmonitored (and significant) areas located below the fall line of the major tributaries. As a part of any expansions of the network, resources should be sought, and approaches developed, to adequately quantify these currently un-monitored sources.
- ***Focused land use monitoring.*** The expanded network should place particular emphasis on the adequate instrumentation and monitoring on the heavily agricultural drainages of the Eastern Shore in which certain best management practices are being emphasized. With regard to urban sources in the watershed, steps should be taken to ensure that more sites that are designed to detect urban nonpoint-source signals from major metro centers near the Bay (Baltimore-Washington-Richmond corridor and the Hampton Roads area).
- ***Network redundancies.*** As plans to expand the monitoring network proceed, steps should be taken to conserve resources by conducting a critical evaluation for redundancies in the existing network. For example, a cursory review of the current network of 33 sites revealed a few sites with similarities in drainage area on the same tributary. Where there is no intervening major source or land use change in the intervening drainage area, some assessment of the utility of similar sites should be conducted.
- ***Addition of small drainage area stations.*** The CBP plan to expand the NTWQMN includes the addition of a large number of monitoring stations on streams draining much smaller watersheds than those currently included in the network. Smaller drainage areas (with the appropriate monitoring approaches) will make it possible to better observe or resolve anthropogenic effects (through stronger signals) than is currently possible in the large basin network. Monitoring in smaller (more homogeneous) drainages will allow simpler, stronger signals from watershed activity to be detected, yet they may vary in their responses

because of landscape and hydrological differences. Careful selection will also allow effects of differences in land use patterns to be discerned. As more small drainages are instrumented, however, the monitoring plan(s) and load estimation methodologies should be closely examined to insure that shifts to smaller watersheds do not render the existing protocols (*e.g.* instantaneous samples, ESTIMATOR) less useful in predicting loads.

- ***Point-source monitoring.*** In the course of this Task Force's discussion with CBP analysts, concerns were raised about uncertainties in annual loading estimates from point source discharges into the Chesapeake Bay, tidal tributaries or watershed rivers. The quality of the point source load estimates and the underlying data reported by wastewater treatment plant operators should be evaluated and quantified. One possibility is that the data variability (and load certainty) is related to the NPDES reporting requirements for wastewater treatment plants (WWTPs) of different sizes. It is quite likely that the large plants, which account for most of the annual WWTP nutrient loads, provide process-monitoring data from daily flow-weighted composite samples, and are, as a result, very-well characterized. Smaller plants may have different monitoring requirements, and as a result, greater uncertainty. As the nontidal monitoring network is expanded, however, care should be taken to ensure that the accuracy of the load estimates from small WWTPs is improved, particularly where they are discharging to smaller streams included in the expanded network. Overall, we believe that steps should be taken to ensure that the point-source loads have the least uncertainty of any of the loading components in the CBP progress assessments and associated models.
- ***Other issues***. As the expanded network design takes shape, care should be exercised to ensure that stations are selected that will be responsive to monitoring needs relevant to other activities in the CBP (*e.g.*, tributary teams). It is also recognized that it will be difficult (but critical) to define and maintain the correct balance (in a universe of limited resources for monitoring) between adding representative vs. targeted sites to the network. The CBP should also differentiate between monitoring sites that are required to document water, nutrient, and sediment loads to the Bay from monitoring sites that are chosen to improve our understanding of hydrological and nutrient transport processes that can improve the WM and our understanding of trends. Monitoring stations established to improve our understanding of watershed processes in small basins, urban streams, and agricultural watersheds do not necessarily have to become permanent parts of the monitoring network.

Presentation of Results and Uncertainties to the Public

There is a clear need to improve the presentation of results from the nontidal water quality monitoring to the public (and inevitably to the media). To date, the scientific and technical community has not enjoyed conspicuous success in conveying key points about the range of natural processes that affect temporal variability in water quality and water quality trends, or the uncertainty associated with the observations. Directed efforts should be undertaken to ensure that publicly disseminated data presentations and trend analyses always address some key and repeating issues:

- Climate variability alone causes huge swings in flow and nutrient loads. Managers need to understand this and communicate to the public just how big these swings have been and can be.
- Effects of improved management are difficult to separate from the high variability of watershed responses, and the evaluation of restoration success is often complicated by the "noise," which makes trend detection difficult.
- Improvement trends will be multi-year (long-term) phenomena. Demonstration of this will require repeated illustrations that natural processes can account for apparent short-term trends—good and bad. It is as important to resist the temptation to claim progress as a result of a single year or short duration with low nutrient loadings (2002) as it is to avoid decrying ineffectiveness when a very wet year (2003) brings high nutrient loadings.
- Lag time is a real phenomenon and a means of meaningfully illustrating it to the public should be found. It is important to demonstrate the degree to which stored nutrient pools in the watershed are changing and to project how these changes might affect future in-stream monitoring observations and loads to the Bay.
- Changes in how data analyses are conducted or reported (e.g., discontinuance in reporting of trend analyses of annual loads or FWC) should be undertaken with care and be fully explained to ensure that the public and media, as well as managers, are well-informed on the reasoning behind any changes.
- Communications regarding the local water quality benefits of CBP nutrient reduction activities should be improved, particularly to stakeholders in more remote parts of the watershed. Analysis and interpretation of data from the expanded monitoring network might be helpful in this regard.
- While not specifically a part of the Non-tidal Water Quality Monitoring Network, we restate our earlier observation that the process of WM formulation, testing and validation should be made more transparent in order to allay concerns in both the technical and lay arenas. As discussed in the next section, more extensive comparisons between nontidal monitoring data and model projections might be helpful in this regard.

Review Processes

The Task Force believes that it is essential that the nontidal monitoring network and program objectives should be periodically reviewed by experts within, and external to, the Chesapeake Bay Program community. Reviewers with expertise in hydrology, water quality, and statistics should be identified and tasked with conducting a critical assessment of existing (and proposed) data analysis methodologies. Major presentations to the public should routinely be subjected to a peer-review process. Further, the notion of periodic independent review should be incorporated into the ongoing nontidal monitoring network operation.

Integration of Monitoring and Modeling

As discussed earlier, monitoring and modeling play distinct and critical roles in support of the effort to reduce nutrient and sediment inputs to the Chesapeake Bay. The primary purpose of the CBP nontidal water quality monitoring is to systematically characterize rates of nutrient and sediment loads from the various river basins to the Bay. At a minimum, monitoring data provide estimates of instantaneous nutrient and sediment loading rates and average loading rates per unit of watershed area to the specific monitoring locations and at specific times of sampling. At present, these data thus provide the best assessment of the annual loadings of constituents from the different basins to the Bay. Additionally, monitoring over longer time intervals (many years) provides an indication of trends in nutrient loading rates. In short, monitoring tells us what is happening, and eventually, what has happened over time. Monitoring in large basins (such as at the RIM stations) tells us little about the factors within the basin that are controlling nutrient loading rates, and thus monitoring cannot predict future nutrient loading rates as activities in the watershed change.

The Chesapeake Bay Watershed Model

The Chesapeake Bay Program's Watershed Model (WM) is refined from the general Hydrologic Simulation Program-Fortran (HPSF) model (Linker et al., 2002) and is now in Phase V development. The WM is linked to a Regional Acid Deposition Model (RADM) to estimate nutrient inputs from atmospheric deposition. The WM and the Chesapeake Bay Estuary Model (CBEM) were used in tandem to set and allocate the CBP nutrient and sediment load goals needed to meet the water quality criteria called for in the Chesapeake 2000 Agreement (Koroncai et al., 2003). In addition, the U.S. Geological Survey has developed and refined a Spatially Referenced Regressions on Watershed (SPARROW) Attributes model, which uses a nonlinear regression approach to spatially relate nutrient sources and watershed characteristics to annual nutrient loads of streams throughout the Chesapeake Bay watershed (Brakebill and Preston, 2001). This is a "hybrid" modeling approach directly incorporating observations from monitoring, which is useful in spatial representation of nutrient sources, although it is not designed to account for temporal changes or to project future scenarios, as is the WM.

The purpose of the WM is to organize what we know about the factors and processes that control flow, sediment movement, and nutrient transport rates into a set of mathematical relationships and linkages that collectively provide an estimate of sediment and nutrient loading rates from a watershed. Modeling, by default, is a simulation of reality and provides an approximation of the system that considers the physical, chemical, and biological processes that control transport rates, and how these processes are affected by external forces such as weather and changes in human activities. Modeling provides a tool for approximating the relative effects of various activities in a watershed on sediment and nutrient transport rates and how future loads could be changed through management activities. Thus, modeling provides a predictive (and speculative) tool to allow investigation of the response of the watershed system to various "what if" scenarios that may be of interest. In short, if the model provides a reasonably correct representation of the system then modeling helps us to understand why nutrient loads are what they are and how they will be changed due to management efforts or major changes in weather, population, landscape or management practices.

Paradoxically, models often yield relatively inaccurate predictions (due to model inadequacy and parameter uncertainty) in cases where model predictions are directly compared to observations. This arises in part because of—not in spite of—model complexity. Ecosystem models require calibration, and can only be calibrated to the extent supported by available data (acquired through some type of monitoring program). Because monitoring estimates are closely linked to observed data and employ parsimonious models (if any), they are viewed to be accurate assessments of conditions and the magnitude of errors (uncertainty) associated with these estimates can typically be estimated statistically. And, for monitoring, uncertainty can be reduced by more frequent observations, improved sampling techniques, and more spatial locations where monitoring is conducted. The problems that arise in watershed modeling and monitoring also arise in the monitoring and modeling used to support weather forecasting, national economic policy, and many other complex systems for which the Nation needs to make decisions.

It is tempting to compare the monitoring estimates against model predictions for times and locations where both are available for a watershed or basin. In conducting such comparisons at any time and location it is necessary to consider three vectors:

- $S(t)$ The true (and unknown) "state of the river" at time *t*. $S(t)$ would comprise the effect of an infinite collection of variables related to chemical, hydrological, biological, and physical characteristics of the river. However, here we will focus only on a small subset of variables associated with nutrient loads entering the Bay.

- $\hat{S}(t)$ The estimated “state” based on monitoring data augmented (mostly interpolated) by statistical methods which automatically quantify the accuracy of estimates and generally avoid introducing substantial bias on them.

- $\tilde{S}(t)$ The estimated “state” based on a process model designed to describe the relationship between forcing variables (such as weather, land use, management practices, etc.) and response. Typically, there is considerable uncertainty in model input parameters and the model itself, and it is not easy to quantify the uncertainty in model predictions.

If the monitoring and process models had no errors associated with them, then one would expect that:

$$S(t) = \hat{S}(t) = \tilde{S}(t)$$

However, the reality is that

$$S(t) = \hat{S}(t) + error_{1,t} = \tilde{S}(t) + error_{2,t}$$

where the error associated with the monitoring can be approximated (based on the statistical analyses associated with the observations), but the error associated with the modeling is difficult to determine and often not available. In the case of the CBP, a primary goal should be to reduce the size of the error associated with both monitoring (through increased and improved monitoring efforts) and modeling (by improving model representation and assessing modeling uncertainty). That is, however, not the only important criterion to consider. For example, it may be more important to describe the impact of a forcing variable, such as climate change or management implementation, on basin response than it would be to estimate the absolute loads.

Using Monitoring Results to Improve Model Performance

Monitoring is critical for model calibration and for identifying model inadequacies and lack of model fit. Thus, the modeling effort must initially rely heavily on monitoring data, and model and monitoring results are frequently compared. It is critical to understand the limitations of using monitoring data to support modeling efforts. In conducting such comparisons it is important that the uncertainties associated with both the monitoring (i.e. how close are monitoring data estimates to "real" flows and loads) and the modeling (what are the model uncertainties as affected by input parameter uncertainties and model structure) be assessed and documented.

In the case of Chesapeake Bay, the River Input Monitoring (RIM) network was established to capture the maximum area of the Bay watershed with a minimum of flow measurement and sampling effort. Whereas data from this network provide an assessment of nutrient loads from a large fraction of the watershed passing through the specific monitoring location at the sampling times, they provide little insight into the factors controlling the observed nutrient loads coming from these river basins. Thus, monitoring results from the RIM network, when considering the uncertainties associated with both the model and the monitoring data, can be used to provide a basic indication of the overall predictive ability of the watershed model at specific points (in time and space). However, such comparisons are not useful for correcting model flaws beyond telling us that something needs to be fixed. For example, the current watershed model does not account for long nitrogen residence times in groundwater that are known to occur in some regions of the watershed, although estimates of these lags are being incorporated into the Phase V WM. Monitoring results are highly influenced in some regions by past watershed conditions, whereas model output is based solely on current conditions, as best represented by the model system. The results should not be expected to be the same unless watershed conditions are static. Monitoring activities that should be useful for improving modeling calibrations and performance include:

- monitoring at smaller spatial scales to isolate more specific land use and hydrogeomorphic conditions;
- monitoring over longer time periods so as to better represent changing conditions;
- increasing sampling frequency so as to better represent changing conditions with time and hydrologic regime, and

- monitoring conditions below the RIM stations (particularly in the Eastern Shore and the urbananizing centers along the Bay).

Using Model Results to Improve Interpretation of Monitoring

Ideally, a watershed model is a mathematical compilation of our collective knowledge of factors and processes controlling flow and constituent transport. If this is the case, then models should contain an explanation for results observed from monitoring efforts. One of the most basic yet valuable results of modeling is that it requires tracking of watershed information that we believe to be critical in controlling flows and nutrient losses. A model should provide insight to such basic questions as: What is the distribution of nutrient sources within the watershed? Where are the sinks for nutrients in the watershed? How have sinks and sources changed over time? Where in the watershed do we think nutrient transport rates have changed? In large watersheds the tracking of basic, but critical information needed to run models is valuable for understanding monitoring results, even though there always are uncertainties regarding how this information is used in the model. Models also should provide guidance as to how monitoring strategies should be adapted to better characterize and understand nutrient transport processes in the watershed. Essentially, this is an ongoing calibration effort.

Another key role of models is that they can be used to estimate nutrient loads from unmonitored parts of the watershed. Unfortunately, in the Bay watershed the primary unmonitored regions are within the Coastal Plain, where residence times of nitrogen in groundwater tend to be longer, and in highly populated, high density urban areas, for which the model algorithms may not be fully verified or representative. Because the current watershed model does not account for temporary storage of nitrogen in shallow groundwater, its use for accurately predicting short-term changes in nitrogen loads as changes in land use occur will be especially limited in much of the unmonitored region of the watershed. Additional effort has to be invested in monitoring in areas below the RIM stations (so that an accurate assessment of such contributions to the Bay can be made) or the model has to be altered to account for the temporal patterns of nitrogen movement through subsurface flow in these systems (and will necessitate comparison to monitoring data so as to test the reliability of such alterations). Similarly, for the major urban areas immediately adjacent to the Bay, model predictions may not be adequately representing loading contributions. Thus, additional monitoring and/or model attention is necessary for these areas as well.

Integrating Monitoring and Modeling to Assess Progress

The current structure of the WM is best suited for developing strategies for reducing nutrient loading rates to the Bay and for predicting possible impacts of major shifts in land use/management or forcing functions on nutrient loading rates. The predictive power of the model is only as accurate as our understanding and model representation of all the factors and processes that combine to determine nutrient transport rates. Because information validating the effectiveness of many of the practices implemented is sparse, input parameters are highly uncertain and the model has large spatial clumping, it is unreasonable to attach a high degree of certainty to model predictions of future nutrient loads. Yet, model results are generally presented with a sense of deterministic accuracy, or at least that is the impression of many managers and the public. A critical present need is to conduct an uncertainty assessment of the model and monitoring results to improve the general understanding of the degree of uncertainty associated with model predictions and the possible reasons for disagreement between model and monitoring results.

There is little choice other than the use of modeling for looking into the future either for planning or predictive purposes. Monitoring will be the ultimate tool for assessing how nutrient loads have responded to management efforts but definitive conclusions will only be achieved in long time frames (decadal or longer), and after the fact (post implementation of management or policy changes and conditions). The use of modeling is especially valuable in areas of the watershed where groundwater flow plays a major role in nitrogen transport. In these areas, monitoring of baseflow nitrogen loads provides information on past land use activities and the effect of current practices can only be assessed with some sort of model. It is unfortunate that the Bay watershed model has not been designed to assess nitrogen transport through groundwater flow systems because this is a function that monitoring cannot address in the short term. Likewise, assessing loads from unmonitored regions of the watershed (assuming that monitoring will not be conducted at all locations) also must be done using some sort of modeling approach. In the long term, monitoring and modeling need to be coupled in an adaptive, iterative process that can lead to model improvements and also to guide monitoring efforts so as to address the greatest areas of uncertainty associated with the model.

Institutional Adjustments for More Effective Integration

The first step in making modeling and monitoring efforts more effective is a critical evaluation and documentation of the shortcomings of both efforts. This should include efforts to clearly document the uncertainties associated with both the monitoring data and the modeling efforts. The monitoring effort in its narrowest sense (the measurement of flows and concentrations at specific locations and times) can stand alone as a tool to assess sediment and nutrient loads to the Bay from a large fraction of the watershed (at present areas above the RIM stations). However, a considerable portion of the contributing area to the Bay lies below the RIM monitoring stations and thus is not included in the “observed” loadings to the Bay. Additionally, it is known that the WM relies on a considerable number of input parameters that are at best crude estimates of actual conditions on the landscape. Also, due to the spatial lumping inherent in the model, considerable over-generalization of the basin occurs in the model representation. As a result, it is expected that a sizeable degree of uncertainty is associated with any model predictions. These modeling uncertainties need to be thoroughly assessed and considered in any use of modeling results.

The definition of monitoring within the Bay program needs to be expanded to include the assessment and collection of various watershed parameters that affect flow and nutrient pools and transport so that overall nutrient transport potential in the watershed can be tracked more systematically. For example, are land use patterns tracked systematically and updated regularly? Are the latest agricultural statistics used for model input? Are there data sets that would make model output more accurate that are not being collected? While the precise effect of certain changes in the watershed on nutrient transport may not be known, the model is the logical place to track those changes. Modeling can serve as a useful tool for guiding such data collection. More effective institutional networks (e.g. between the modeling and monitoring subcommittees of the Implementation Committee) need to be established that ensure that the necessary data sets are updated regularly and provided in a usable format to the modeling unit at the CBP.

Modeling is also a useful tool for identifying research needs with regard to the processes that affect nutrient delivery rates. Watershed processes that attenuate nutrient losses currently are one of the most poorly understood aspects of watershed nutrient transport. This applies both to promoted practices such as riparian buffers and nutrient management planning, as well as natural attenuation mechanisms. River monitoring data provide an integration of nutrient sources and sinks within a watershed, but provides little information useful for calibrating the specific sink and source functions within a model. Improving how the model addresses nutrient attenuation within a watershed will require interaction with the research community to ensure that the approach used is consistent with current research. In some cases information is very incomplete and model shortcomings can be used for establishing research priorities. Because the Bay watershed is so large and diverse, the most effective strategy may be to establish research teams that evaluate model functions in specific sub basins or regions of the watershed. Overall there appears to be a need for more rigorous technical oversight of the modeling effort by practicing research scientists on an ongoing basis. To date, model calibration has been focused at the large scale but more emphasis should be placed on calibration at smaller spatial scales and for specific land uses. This will require communication between modelers and scientists and managers conducting small-scale research and monitoring projects.

Comparing Model Estimates and Monitoring Trends

As noted above, an assessment of the uncertainties associated with both monitored data and model predictions is critically needed. Such uncertainty assessments would thus allow some sense of the correspondence that could reasonably be expected when comparing modeling simulations to observations. Additionally, assessing input parameter uncertainties would allow a strategic focusing of more in-depth quantification and collection of appropriate model input data as well as improved understanding of the reasonableness of simulated outcomes from various modeling scenarios.

- Some assessments should be made of WM prediction uncertainties. For instance, are model predictions within + or – 50%, 100% etc.? Also, which model predictions are most uncertain and why? What model processes are least well represented? And, explanations should be provided of what these uncertainties mean relative to any decision-making.
- Model output should be presented in the framework of the factors that generate base loads and what specific change in input parameters caused projected changes in nutrient loads. It should be clear as to why the model indicated a change (or

lack of change) in nutrient loads. This will help move beyond the model as a black box, and should provide more insight as to why modeling and monitoring results differ. For example, if model projected reductions in nitrogen loads were broken down into leaching and overland components it would be readily apparent that delivered load reductions should not be expected to match up with monitoring data on a year by year basis.

Specific basins or subbasins should be modeled individually with model performance reviewed and assessed by more locally based and discipline-specific scientific personnel. This would allow a more regional representation of the model formulation and evaluation. Monitoring data, at both basin-scale and more localized (where available), could be used to assess how well the model represents reality.

Addressing Differences in the *Factors Affecting Nutrient Trends* Report

Analysts within the USGS and EPA are presently undertaking an update of the earlier report by Sprague et al. (2000), which was an attempt to bring together the available models and monitoring data to assess the factors affecting nutrient trends within the watershed. There can be many reasons for discrepancies between monitored and modeled gains or reductions for the major river basins draining to the Bay. One must assume that the monitored data provide the most correct representation of what is actually occurring for any given time frame analyzed, but there is some uncertainty associated with these measured loads, especially for storm flow from smaller basins. The model, on the other hand, represents a large basin as a collection of relatively few large “lumped” parameter subbasins, which thereby represents only a crude representation of the system. Input parameters that are used to describe the system state are generally estimated from secondary data, again in a relatively simplistic manner. Additionally, the complexity of generation and transport of constituents that occurs across the landscape likely will not be correctly portrayed within the model. As various scenarios are modeled the true processes and pathways of generation and transport may not be correct. All of this leads to considerable uncertainty in both the absolute magnitudes of flow and load predictions. Thus, it is not unexpected that monitored and modeled scenarios may show unexpected differences; and in fact such differences may simply be within the bounds of “uncertainty” associated with the monitoring and modeling.

Although the lack of high quality input data sets and major information gaps regarding watershed nutrient and sediment transport and attenuation processes limit model predictive ability, the first step to dealing with “unexplained” changes in watershed nutrient loads is to identify the factors that are driving nutrient loads within the model. In some cases there may be simple explanations for why model output does not match monitoring data (i.e., subsurface nitrate storage), but in other cases it may be that an important mechanism is misrepresented or that there is too little scientific information for adequate characterization. Yet in other cases it may be that something is changing in the watershed that is not being captured by input data sets for the model. Basically, there needs to be an ongoing effort to identify and correct model deficiencies as new scientific results and techniques for collection of pertinent data sets emerge. The critical step in this process is to systematically link modeled nutrient loads to the factors that drive total loads; more specifically, there will be a greater likelihood for sorting out "unexplained" loads if how the model generates and attenuates nutrient loads within a watershed is clearly presented.

Addressing the Unmonitored Parts of the Bay Watershed

It is estimated that approximately 24% of the TN and 30% of the TP Bay loadings are from unmonitored nonpoint sources below the RIM stations (Gary Shenk, personal communication). These unmonitored areas include high-density and rapidly developing urban areas and near-Bay agricultural areas including the Eastern Shore. Because of the proximity of these areas to the Bay, this region is a prime target for cost-effective nutrient reduction practices. Efforts should be directed at the following:

Eastern Shore:

- Additional monitoring strategies that allow assessments of contributions to the Bay from Eastern Shore watersheds are critically needed.
- Identify any observed data on flows and loads for watershed areas on the Eastern Shore (or similar Coastal Plain watersheds) and use these data as a relative check against the predictions of loads from SPARROW used in the WM or other estimation approaches.
- Utilizing an expert scientific panel, evaluate the loading predictions for Eastern Shore watershed scenarios and the reasonableness of model representations and obtain recommendations of how model simulations can be more representative of present and future loadings to the Bay.
- Use a smaller-scale and more process-based model (which will better incorporate lag time issues) to predict loadings from Eastern Shore conditions and then compare these with the estimations used in the WM. [Note that this does introduce the issue of "scale of lumping" as one more confounding problem.] Also, one should use available observations from local watersheds as a check against the other model predictions.

Urban areas:

- Identify any observed data on flows and loads for comparable urban areas and use these data as a relative check against the predictions of loads from SPARROW used in the WM or other estimation approaches. Check, in particular, the nutrient yield estimates that have been developed from the Baltimore Ecosystem Study.
- Use an expert panel to evaluate the reasonableness of the predictions for unmonitored urban watersheds and have the panel recommend modifications to the procedures so as to better represent the urban area contributions or at least estimate errors associated with such predictions.
- Identify, if possible, better urban stormwater models that have been better verified for urban areas and compare the WM predictions against these predictions to

assess relative performance. Assess the magnitudes of the variations and discrepancies.

Modeling Specific Years and Conditions

The spatial distribution of precipitation is one of the most important variables in watershed hydrologic modeling. NOAA has implemented a program that produces precipitation estimates at approximately 5-km resolution at 6-hour intervals based remote sensing involving both NEXRAD radar scanning and satellite imagery. With such inputs, it should be possible to run model simulations for hydrologic scenarios very close to actual conditions. Techniques for remote sensing of other watershed factors that influence nutrient loads also have advanced dramatically since the Bay restoration effort began. This creates the opportunity for comparing modeled nutrient loads to those calculated from monitoring data under very similar hydrologic conditions. Annual runs with actual precipitation inputs create opportunities for evaluating the model, and also make the model more useful for estimating in a given year where nutrient loads are generated within the watershed. At large watershed scales, annual simulations still will not eliminate all of the uncertainty regarding the relative magnitude of sources and sinks. This can only be achieved through modeling at smaller basin scales and in settings where individual land uses and nutrient control practices can be isolated. But for the model to have credibility for planning and predictive purposes, it should have the capability of producing annual loads that are similar to those calculated from monitoring data. As stressed earlier, this should be an iterative process as new research findings and data collection techniques emerge. Ultimately, model capability is limited by our understanding of watershed nutrient retention and transport processes and our ability to identify and assemble accurate input data sets. Model shortcomings are useful for identifying research and data needs, and annual simulations should speed up the process of identifying these needs.

References

Band, L., K. Campbell, R. Kinerson, K. Rechhow, and C. Welty. 2005. *Review of the Chesapeake Bay Watershed Modeling Effort*. Chesapeake Research Consortium, Edgewater, MD.

Boesch, D.F. 2000. Measuring the health of the Chesapeake Bay: toward integration and prediction. *Environmental Health Research* 82: 134-142.

Boesch, D.F. 2004. Adaptive management could create a model monitoring approach. *Bay Journal* 14(8).

Brakebill, J.W., and S.D. Preston. 2001. *Digital Data Used to Relate Nutrient Inputs to Water Quality in the Chesapeake Bay Watershed, Version 3.0*. Open-File Report 2004-1433**.** U.S. Geological Survey, Baltimore, MD.

Chesapeake Bay Program. 1999. *Environmental Outcome-Based Management: Using Environmental Goals and Measures in the Chesapeake Bay Program*. CBP/TRS 248/00, EPA 903-R-00-016. U.S. Environmental Protection Agency, Annapolis, MD.

Chesapeake Executive Council. 1999. *Chesapeake 2000: A Watershed Partnership*. Chesapeake Bay Program, Annapolis, MD.

Chesapeake Bay Program, 2002. *The State of the Bay: Report to the Citizens of the Bay Region*. CBP/TRS 260/02, EPA 903-R-02-002. U.S. Environmental Protection Agency, Annapolis, MD.

Chesapeake Bay Program. 2004. *The State of the Chesapeake Bay and Its Watershed*. CBP/TRS 273/05, EPA 903-R-04-009. U.S. Environmental Protection Agency, Annapolis, MD.

Chesapeake Bay Watershed Blue Ribbon Finance Panel. 2004. *Saving a National Treasure: Financing the Cleanup of the Chesapeake Bay*. U.S. Environmental Protection Agency, Annapolis, MD.

Chesapeake Watershed Partners. 2004. *Memorandum of Understanding Regarding Cooperative Efforts for Monitoring and Assessing Water Quality in the Streams and Rivers of the Chesapeake Bay Watershed*. Chesapeake Bay Program, Annapolis, MD.

Cohn, T.A., D.A. Caulder, E.J. Gilroy, L.D. Zynjuk, and R.M. Summers. 1992. The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay. *Water Resources Research* 28:2353-2363.

Duckstein, L., S. Ambrus, and D.R. Davis. 1985. Management forecasting requirements. In: *Hydrological Forecasting* (M.G. Anderson and T.P. Burt, Eds.), Chap. 16, pp. 559-585, John Wiley and Sons, New York.

Ernst, H.R. 2003. *Chesapeake Bay Blues: Science, Politics, and the Struggle to Save the Bay*. Rowman and Littlefield Publishers, Inc., Lanham, MD, 203 pp.

Koroncai, R., L. Linker, J. Sweeney, and R. Batiuk. 2003. *Setting and Allocating the Chesapeake Bay Nutrient and Sediment Loads: The Collaborative Process, Technical Tools, and Innovative Approaches*. EPA 903-R-03-007. U.S. Environmental Protection Agency, Annapolis, MD.

Langland, M.J., J.D. Blomquist, L.A. Sprague, and R.E. Edwards. 1999. *Trends and Status of Flow, Nutrients, and Sediments for Selected Nontidal Sites in the Chesapeake Bay*. Open-File Report 99-451. U.S. Geological Survey, Reston, VA.

Langland, M.J., S.W. Phillips, J.R. Raffensperger and D. Moyer. 2004. *Changes in Streamflow and Water Quality in Selected Nontidal Sties in the Chesapeake Bay Basin, 1985-2003*. Scientific Investigations Report 2004-5259. U.S. Geological Survey, Reston, VA.

Linker, L.C., G.W. Shenk, P. Wang, K.J. Hopkins, and S. Pokharel. 2002. A short history of Chesapeake Bay modeling and the next generation of watershed and estuarine models. *Water Environment Federation Watersheds 2000 Conference.*

National Research Council. 2004. *Adaptive Management for Water Resources Project Planning*. National Academies Press, Washington, DC.

Preston, S.D., and J.W. Brakebill. 1999. *Application of Spatially Referenced Regression Modeling for the Evaluation of Total Nitrogen Loading in the Chesapeake Bay Watershed*. USGS Water-Resources Investigations Report 99-4054, U.S. Geological Survey, Baltimore, MD, 12 pp.

Scientific and Technical Advisory Committee. 1997. *Integrated Analysis of Chesapeake Bay Monitoring Data*. STAC Publication . Chesapeake Research Consortium, Edgewater, MD.

Scientific and Technical Advisory Committee. 1998. *Watershed Response to Changes in Nutrient Loads: The Best Uses of Monitoring and Modeling*. STAC Publication 98-1. Chesapeake Research Consortium, Edgewater, MD.

Scientific and Technical Advisory Committee. 2005. *Assessing the Spatial Representativeness of the Chesapeake Bay Non-tidal Water Quality Monitoring Network.* STAC Publication 05-006. Chesapeake Research Consortium, Edgewater, MD.

Sprague, L.A., M.J. Langland, S.E. Yochum, R.E. Edwards, J.D. Blomquist, S.W. Phillips, G.W. Shenk and S.D. Preston. 2000. *Factors Affecting Nutrient Trends in Major Rivers of the Chesapeake Bay Watershed*. Water-Resources Investigations Report 00-4218. U.S. Geological Survey, Richmond, VA.

U.S. Commission on Ocean Policy, 2004. *An Ocean Blueprint for the 21st Century*. Final Report of the U.S. Commission on Ocean Policy, Washington, D.C.

Whoriskey, P. 2004. Bay pollution progress overstated. *The Washington Post*, July 18, 2004, p. A.01.