VIRGINIA WATER RESOURCES RESEARCH CENTER

REPORT OF THE WATER QUALITY ACADEMIC ADVISORY COMMITTEE

SPECIAL REPORT

Virginia Tech

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY
BLACKSBURG, VIRGINIA

1998

Report of the Water Quality Academic Advisory Committee

Prepared for the

Virginia Department of Environmental Quality, Division of Water Program Coordination
Assessment and Planning

Leonard A. Shabman (Chair)
Virginia Water Resources Research Center
Virginia Tech

Carl Hershner
Resource Management & Policy
Virginia Institute of Marine Science

Howard I. Kator
School of Marine Science
Virginia Institute of Marine Science

Eric P. Smith
Department of Statistics
Virginia Tech

Leonard A. Smock
Biology Department
Virginia Commonwealth University

Tamim Younos
Virginia Water Resources Research Center
Virginia Tech

Shaw L. Yu
Civil Engineering & Applied Science
University of Virginia

Carl E. Zipper
Crop and Soil Environmental Science
Virginia Tech

Virginia Water Resources Research Center
10 Sandy Hall
Virginia Tech
Blacksburg, VA 24061

October, 1998

VWRRC Special Report No. SR8-1998

25711

This special report is a publication of the Virginia Water Resources Research Center, and was published with funds provided in part by the Virginia Department of Environmental Quality. The contents of this publication do not necessarily reflect the views or policies of the Virginia Department of Environmental Quality or the Virginia Water Resources Research Center. The mention of commercial products, trade names, or services does not constitute an endorsement or recommendation.

Additional copies are available while the supply lasts and may be obtained from the
Virginia Water Resources Research Center
10 Sandy Hall
Blacksburg, VA 24061
(540)231-5624
FAX: (540)231-6673
e-mail: water@vt.edu
homepage address: http://www.vwrrc.vt.edu/vwrrc/vwrrc.htm

Single copies are free to Virginia residents.

Leonard A. Shabman, Director

Summary

The Virginia General Assembly in the 1997 *Water Quality Monitoring, Information and Restoration Act (WQMIR)* directed the DEQ to develop the EPA-required 303(d) and 305(b) reports in consultation with experts from the states' universities. Also, WQMIR requires the DEQ to "develop and publish a procedure governing its process for defining and determining impaired waters and shall provide for public comment on the procedure" with the assumption that these 303(d) procedures will be developed after consultation with scientists from the states' universities.

To meet the WQMIR academic consultation requirements, the DEQ asked the Virginia Water Resources Research Center (VWRRC) to organize and coordinate a Water Quality Academic Advisory Committee (WQAAC) to serve as an independent advisory body to the DEQ. The WQAAC reviewed and evaluated the scientific merits of the DEQ's existing and evolving water quality assessment procedures for the 305(b) and 303 (d) reports. Some of the WQAAC recommendations were suggestions for action that result in immediate improvements in the assessment reports. In addition, the WQAAC suggested actions that would lead to improvements in the assessments in the long term.

Overall the WQAAC made 17 findings and recommendations to the DEQ. Each of these should be considered separately. In making these findings and recommendations, the WQAAC recognized that the assessment process is a work in progress and that there are few "correct" ways to approach the assessment challenge. In addition the WQAAC was sensitive to the reality that assessment rests on both science and on policy judgements.

As an overall summary, the WQAAC did not recommend that DEQ make immediate changes to its assessment guidelines. However, the WQAAC did recommend a number of future actions. These are summarized as follows:

- that the DEQ carefully explain the logic of its assessment approaches and the logic employed to list waters as impaired by organizing its arguments. This might be accomplished by using concepts from the risk assessment literature and by improved explanations of current use of statistical inference procedures.

- that three new Technical Advisory Committees (TAC) be formed to address pressing issues and to help the DEQ to make future modifications to the guidelines and assessment practices. These TACs would focus on nutrient enrichment criteria, on sediment contamination, and on monitoring strategy. Long term strategies to further develop the monitoring program were suggested by the WQAAC and might be considered by the monitoring TAC.

- that the DEQ conduct a review of certain practices to determine whether long term adjustments are warranted. Such review was suggested for the criteria for shellfish contamination, bio-monitoring, and model development and use for the TMDL process.

- that based on a review of the current monitoring program and its costs, the DEQ should find ways to redesign its program to increase sampling frequency and locations while limiting costs. Costs might be limited by increased cooperative work with others. From this effort, the DEQ would enhance its ability to define the geographic extent of impairment, categorize long term trends in water quality, and support the modeling requirements for the TMDL process.

TABLE OF CONTENTS

Report of the Water Quality Academic Advisory Committee

I. Introduction

The Federal Water Pollution Control Act (PL92-500), commonly known as the Clean Water Act, last authorized by the Water Quality Act of 1987 (PL100-4), establishes a process for States to develop information on the quality of their water resources and to report the information to the U.S. Environmental Protection Agency (EPA), to the U.S. Congress, and to citizens. The procedures governing this process are extracted from Sections 106(e), 204(a), 303(d), 305(b), and 314(a) of the Clean Water Act.

The Virginia Department of Environmental Quality (DEQ) is required by the EPA to prepare a biennial 305(b) water quality assessment report describing the general condition of the waters of the state. The 303(d), Total Maximum Daily Load (TMDL), report describes the condition of particular waterways, and specifically identifies waters deemed to be "impaired" so that remedial actions should be taken to achieve water quality standards. Both reports are submitted to EPA on April 1 of even numbered years.

A 1996 Joint Legislative Audit and Review Committee (JLARC) report on the Virginia Department of Environmental Quality (DEQ) and a Chesapeake Bay Foundation (CBF) report "Virginia's Waters Still at Risk" were critical of the 303(b) and 305(d) reports prepared by the DEQ. Some of the concerns expressed in JLARC and CBF reports, such as identifying state waters impaired by nutrient over- enrichment, were not included in previous DEQ 303(d) reports because Virginia does not have water quality standards for nutrients. To address other concerns, such as the determination of impaired waters based on monitoring data, the DEQ was directed to review its existing procedures to determine if improvements can be made.

The Virginia General Assembly in the 1997 *Water Quality Monitoring, Information and Restoration Act (WQMIR)* directed the DEQ to develop the EPA-required 303(d) and 305(b) reports in consultation with experts from the states' universities Also, WQMIR requires the DEQ to "develop and publish a procedure governing its process for defining and determining impaired waters and shall provide for public comment on the procedure" with the assumption that these 303(d) procedures will be developed after consultation with scientists from the states' universities.

II. Purpose

To meet the WQMIR academic consultation requirements, the DEQ asked the Virginia Water Resources Research Center (VWRRC) to organize and coordinate a Water Quality Academic Advisory Committee (WQAAC) to serve as an independent advisory body to the DEQ. The responsibility of the WQAAC was to review and evaluate the scientific merits of the DEQ's existing and evolving water quality assessment procedures for the 305(b) and 303 (d) reports. Based on its review, when deemed necessary, the WQAAC made recommendations to the DEQ to modify its assessment guidelines. Some of the WQAAC recommendations were suggestions for action that would result in immediate improvements in the assessment reports. In addition, the WQAAC suggested actions that would lead to improvements in the assessments in the long term. Responsibility for long term actions may rest with the DEQ as well as other agencies and with the General Assembly.

III. Procedure

The WQAAC worked in parallel with the DEQ as the DEQ prepared its reports for submission to the EPA. The WQAAC reviewed the DEQ's proposed guidelines that were published in 1998. The WQAAC agreed to provide feedback on eight priority questions posed by the DEQ. Based on the guidelines review and a meeting held on December 12, 1997 with DEQ staff, the WQAAC identified other relevant issues not specifically identified by DEQ, when the WQAAC commented agreed that they had relevance to the quality and utility of the 305 (b) and 303 (d) reports.

Water quality assessment is as much art as science and is as much policy judgment as it is analysis. However, both the art and the science of water quality assessment are continuously changing. The WQAAC recognized that its advice can be only that – advice -- and that the final determinations of how the assessment program will be conducted and decisions made must rest with the agencies of the Commonwealth. It is also recognized that a state- of the-art monitoring and assessment system can only be developed over an extended period of time with more experience, and perhaps with additional financial resources.

IV. Priority Issues for WQACC Review

The DEQ charge for the WQAAC included eight priority issues. The WQAAC reviewed these priorities and they are described as follows:

1. Use of the Binomial Procedure to Determine Impaired Waters

To characterize impaired waters, for conventional parameters (non-toxic), the DEQ uses EPA's definition of impaired waters, that is a violation rate of greater than 10 percent for numeric water quality standards (referred to as the percentage method). However, many of DEQ's monitoring stations are sampled quarterly and during a 2-year assessment period they yield only 6 to 8 samples. With sample sizes fewer than 10, one sample exceeding the standard will result in the water sample being declared impaired. The DEQ has recognized that the chance of a false classification of a site as impaired is relatively high for this method. To improve the process and directly estimate rates associated with false signals, DEQ in its 1996 report used a binomial probability approach to determine the violation percentage rather than using a simple percentage calculation. Using the binomial method was criticized in the JLARC report. JLARC asserted that the binomial procedure underestimates the violation rate. For example, for a monitoring station that has a data set of 8 samples, using a simple percentage calculation, the water would be listed as impaired if one sample exceeds the water quality standard (violations exceeding 10%). However, for the same sample size, using the binomial method, one sample out of eight in violation would not be deemed unusual, and therefore the water would not be listed as impaired.

25717

2. Criteria for Defining Nutrient Over-enrichment

Virginia does not have water quality standards for nutrient enrichment and therefore, the DEQ did not address nutrient enrichment in the 1994 and 1996 303(d) reports. The absence of nutrient standards is common throughout the nation. The EPA, WQMIR, and the JLARC have suggested that the DEQ include nutrient enrichment in subsequent 303(d) reports. In 1987, a Technical Advisory Committee (TAC) appointed by the Water Control Board recommended a combination of narrative and numerical nutrient standards for lakes, flowing streams, estuaries, and tidal fresh waters. The Attorney General's office advised the agency to adopt a narrative standard to designate nutrient enriched waters. Based on the TAC recommendations, the Chesapeake Bay and its tributaries and Smith Mountain Lake and its tributaries became the two "formally adopted" designated nutrient enriched waters in the state. The DEQ proposes to use the 1987 TAC recommendations as a start to develop procedures for including nutrient over-enrichment in the 303(d) report.

3. Criteria for Defining Shellfish Consumption Restrictions

Waters where the Virginia Department of Health prohibits the direct marketing or depuration of shellfish were listed as impaired in the 1996 303(d) report. Waters that Department of Health has classified as restricted (depuration or relaying of shellfish allowed) were not listed in the report as impaired waters. The WQMIR directs the DEQ to include such "restrictions" as an indicator of impairment in future 303(d) reports. The DEQ proposes to determine the use support for shellfishing based on the determination of restriction on the harvesting and marketability of shellfish resources made by the Virginia Division of Shellfish Sanitation (DSS) as of June 30, 1997. The DSS uses four classifications for describing the status of shellfish waters: approved, conditionally approved, restricted, and prohibited.

4. Contamination of Sediments

Virginia does not have water quality standards for sediment pollution and therefore, it was not included in the DEQ 1996 303(d) report. The WQMIR directs the DEQ to address sediment pollution in future reports. The DEQ proposes to compare the toxics data that have been collected from sediment samples in Virginia to Effect Range-Median (ER-M) values that are used as sediment criteria in the assessment of aquatic life support. The ER-M value for a particular substance represents the fiftieth percentile of all sediment concentration values observed or predicted to adversely affect aquatic biota. The DEQ procedure is based on a 1990 NOAA technical memorandum (The Potential for Biological Effects of Sediment-Sorbed Contaminants Tested in the National Status and Trends Program by Edward A. Long and Lee G. Morgan). In the 305(b) report, the DEQ proposes to use only those values for which the above document reports moderate to high confidence. If there is no ER-M value for a substance, for those substances with sufficient data, the ninety-ninth percentile value from the cumulative frequency distribution for Virginia should be used in determining use support. The ninety-ninth percentile was chosen because, for those substances for which the ER-M and percentile values (from STORET for Virginia) are available, the ninety-ninth percentile values were closest to the ER-M values.

5. Geographically Identify Impaired Waters

The WQMIR directs the DEQ to identify the geographic extent of impairment of stream segments or areas of an estuary. Determining the area-wide extent of the impairment based on the single point monitoring station is dependent upon a number of factors such as land use and potential pollution sources. As a general rule, a single monitoring station should not be used to represent an entire watershed unless land use is relatively homogenous and there is an absence of potential upstream pollution from point and non point sources. In the past, many of the DEQ's Ambient Water Quality Monitoring (AWQM) stations were selected due to point source problems (VPDES permit discharges). In recent years, some stations have been selected to monitor nonpoint source problems. In past 305(b) water quality assessment reports, there has been little consistency between the regions for determining the miles of stream impairment associated with each monitoring station. Most regions have striven to have at least one AWQM station in a watershed. If that station is determined to be representative of that watershed, then the total stream miles associated with that watershed were considered assessed. When an assessment revealed an impairment in water quality then the assessed miles for that specific monitoring station have been limited to a distance upstream and downstream which contains no significant change to water or habitat quality. The remaining stream miles have been evaluated as not assessed. In order to provide consistency between the regions and to get an accurate number of assessed stream miles in the state, the DEQ is proposing the following guidelines: 1) One monitoring station should not be used to assess an entire watershed unless land use, source, and habitat are relatively homogenous; 2) Typically no more than 10 miles of stream should be associated with a monitoring station for conventional pollutants as per EPA guidance. Miles assessed for a toxic pollutant or biological impairment may vary from the miles assessed for conventional pollutants; 3) When determining the miles assessed for a monitoring station, the following items need to be considered: a) point or nonpoint source input to a stream or its tributaries; b) changes in watershed characteristics such as land use; c) changes in riparian vegetation, stream banks, substrate, slope, or channel morphology; d) large tributary or diversion, or; e) hydrologic modifications such as channelization or a dam. The DEQ is recommending that the above approach be phased in over the next couple of 305(b) assessment periods due to the many different considerations which must be made especially for physically or geographically changing watersheds.

6. The 303(d) and 305(b) Reports Shall Identify and Summarize Trends

The WQMIR directs the DEQ to identify and summarize water quality trends in the 303(d) and 305(b) reports. At present, the VWRRC is coordinating a research project in which, researchers are using improved statistical procedures to perform trend analysis for about 200 DEQ monitoring stations. As part of that study, a report will be prepared which will contain summary tables of the parameters analyzed and a watershed scale analysis of trend directions and magnitudes. This is expected to improve DEQ's trend assessment procedures. However, long term trend assessment protocols may still need to be developed using new statistical and data collection protocols.

7. Fully Supporting But Threatened Waters

The EPA and WQMIR direct the DEQ to add a category "fully supporting but threatened waters" to the 303(d) report. In response to EPA's request and public comment, this section was added to the DEQ's 1996 report. Waters for which data evaluation, trend analysis, or other water quality indicators show a possible decline in water quality or a potential water quality problem are placed in this category. Waters are designated as fully supporting but threatened when there is a loss of a designated use documented by ancillary data, such as recurrent fish kills, or documented as polluted by non-agency studies or reports. Threatened waters generally show violations of water quality standards for conventional parameters and require additional monitoring data.

8. Impaired Waters Due to Natural Conditions

The EPA and WQMIR direct the DEQ to add this use category to the 303(d) report. These are waters that are designated as impaired but the impairment is caused by natural conditions and not by past or present human activities. In response to the EPA's request and public comment, this section was added to the 1996 report. However, the State will not implement control measures, fund pollution prevention or reduction projects, or develop TMDLs for these waters.

V. WQAAC Comments on DEQ Priority Issues

1. Use of the Binomial Procedure to Determine Impaired Waters

WQAAC Finding: The Binomial method is a commonly used approach for evaluating whether water-quality standards are exceeded when sample size is limited. This approach differs from taking a simple percentage of the actual samples to determine a violation. The percentage method has a fixed definition of the percentage of measurements required for deciding a site is impaired but the rate of false designations of sites as impaired varies. The Binomial method varies the percentage of sites in violation for a site to be declared impaired but tries to fix the rate of falsely designated impaired sites. The DEQ's use of the binomial method is consistent with standard practice for conventional contaminants. By using the simple percentage approach, it is more likely to conclude that there is an impairment than by using the binomial method. It is also the case that the simple percentage is more prone to drawing a false positive inference that a stream segment is impaired. The DEQ can take actions to improve its application of the binomial method. However, in the long run the DEQ should increase sampling frequency at its monitoring stations. Increasing sampling frequency, i.e., sample size , would make differences between the methods less important and reduce the chance of making a false decision about impairment.

At present, the DEQ - uses the simple percentage approach to - assess - standard violations for toxic contaminants. This difference in approach between types of contaminants reflects an implicit DEQ decision that a false positive is more acceptable and less likely using the percentage method for toxics than for conventional pollutants. This decision is based on a risk management policy judgment.

The Binomial method requires tables or a computer to evaluate the number of violations needed to declare a site impaired. As sample size increases, use of tables is more difficult and an approximation is sometimes used based on the normal distribution. The normal approximation to the binomial distribution is well studied and commonly used in testing and confidence interval situations. The method will provide a reasonable approximation to binomial probabilities when the sample size is large relative to the success (p) or failure (1-p) probability. The rule of thumb is if np and n(1-p) are larger than 5, then the approximation will work well. The main reason for using the approximation is to simplify hand calculations. With the speed of today's computers, it is possible to have the computer do exact calculations, and this is the preferred approach.

When the sample size is small, the binomial method (as proposed) does not have much chance as the percentage method of detecting a site not in compliance. However, we do believe that the binomial approach is the statistically appropriate way to assess risk when there are small sample sizes, and we would recommend its use over the simple percentage as long as sample sizes are small. With smaller sample sizes, the DEQ should consider the chances of making false decisions when selecting a criterion for impairment (e.g the number of samples in violation). Both the chance of a false classification as impaired and a false classification as safe need to be considered.

Also, the binomial method does not take into account the actual value or magnitude of an observation but rather only if the measurement exceeds the standard. As a result, this approach ignores the information about the magnitude of an observation. The DEQ should investigate the use of a method referred to as "acceptance sampling by variables" (Duncan, A.J. 1974. Quality Control and Industrial Statistics R.D. Irwin, Inc. Homewood and Montgomery, D.C. Introduction to Statistical Quality Control. John Wiley, New York). In Duncan's approach the actual magnitude of the observation is taken into account rather than if it is just above or below the standard. The Duncan approach can substantially reduce the sample size without changing the rates associated with false decisions (relative to the binomial method).

The current DEQ approach to standards assessment does not take into account the number of constituents that are sampled. As the number of constituents increase, the probability of observing one exceeding the standard due to chance also increases. An approach to adjust for this problem is to alter the critical value associated with the probability of falsely declaring a site impaired (Type I error rate). Discussion of this problem in an environmental context is in the text by Gibbons (Gibbons, R.D. 1994. Statistical Methods for Groundwater Monitoring. Wiley Interscience, New York.) on groundwater monitoring.

In using the binomial method, error rates need to be evaluated. The problem of error rates is typically viewed in the context of hypothesis testing. The table below illustrates decisions which may be made in the process and two errors result. Both the probability of false rejection (Type I error) and false acceptance (Type II error) need to be considered. Errors can be made in rejecting or not rejecting the null hypothesis. For example, if the null hypothesis is that a particular stressor has no effect, the null hypothesis is not rejected if the variation in data is large relative to the signal in the data.

	Truth	
Decision	Null (no effect)	Alternate (effect)
Null (no effect)	Ok	Type II error
Alternate (effect)	Type I error	Ok

Just because we have not rejected the hypothesis does not imply its truth. The hypothesis may be incorrect but we have not designed the study well enough to detect the signal. The approach that is usually used to help aid in the design of experiments is the power of the test. This is the probability that the hypothesis is rejected when in fact the hypothesis is false. Proper statistical design of studies will focus on the power and choose a sample size to adequately detect important signals. Peterman (1990) has stressed this point in evaluating design of impact assessment programs for power plants. In many studies, the sample sizes that were used were insufficient to assess any but a gross change in ecological conditions. Hence, some of the studies may have indicated no effect due to the power plant when in fact one occurs. Even with adequate power, detection of change or differences may be hindered by improper sampling, natural and unnatural influences and confounding factors. For example, an effluent that is discharged into a rocky substrate may have no effect on the biota because there is little biota there to affect. Factors such as floods may differentially alter habitats in control and impact sites and make them no longer comparable.

Hypotheses about recovery, no effects and safety of stressors are difficult hypotheses to evaluate from a purely hypothesis testing approach. In these studies it is desired that the null hypothesis be not rejected. The above truth table may be represented in terms of environmental effect:

	Truth	
Decision	Site not impaired	Site impaired
Site not impaired (safe)	Ok	Type II error
Site impaired (hazardous)	Type I error	Ok

An example of where this is important is in estimating safe levels of toxicants. One approach is to set up an experiment in which groups of organisms are exposed to different concentrations of a toxicant. Tests are carried out comparing the control (zero dose) to concentrations. If the effect of the concentration is not statistically different from the control, the concentration is said to be safe. Only when the null hypothesis of no difference between control and toxicant concentration is rejected statistically, is the concentration viewed as unsafe. As pointed out by Parkhurst (1990), the burden of proof is placed on the scientist to show that a concentration is more toxic than the control, rather than placing the burden of proof on those who would use the toxicant to show that a dose does not cause an effect. The DEQ should consider evaluating the error rates from a risk management view. A possible approach is to balance the two error rates rather than focus on one error.

It is therefore important that gross uncertainties about hypotheses tests be controlled by proper statistical control of studies. Small sample sizes, poorly designed experiments, ignorance of the proper variables to measure or proper times to sample may lead to acceptance of no effect when in fact there is an effect but high uncertainty. Overall, the DEQ should consider the feasibility of collecting additional information in cases where results are marginal.

An additional insight into the procedure can be gained by considering the question "How small is small sample size? It is difficult to give a definitive number because the useful sample size depends not only on the number of observations but also on the magnitude of violation which is of interest to detect. Here's the problem. The assumption is that there is background "violations" which do not indicate a true violation of standards. Suppose that for some measure the standard is 1.0. In the binomial method, p is the probability that the standard is exceeded in a reference site. Suppose this is 0.25. so that 25% of the samples would be in violation. If a sample of size 4 is taken, we expect to see 1 of the 4 exceeding the standard. The criteria for rejection of the sample (i.e. saying the site is in violation) using the Binomial method is if 2 or more of the measurements exceed the criterion. The value is based on the table below which gives the cumulative probability distribution for a Binomial distribution with p=0.25. We want 1-cumulative probability to be around 0.05 for a standard test. This would occur for 2 violations. Hence we would decide on a violation when we observe more than 2 violation (i.e. 3 or 4).

Table: Cumulative probabilities for Binomial models with different chance of violation and N=4. The cumulative probability is the probability of observing this many or fewer sites in violation. One minus the cumulative probability gives the chance of observing more than this number of violations. Thus when p=0.25, the chance of observing 2 or fewer violations is 0.949. The chance of 3 or more violations is 0.051.

Number of violations	"bin p=.25"	"bin p=.5"	"bin p=.75"
0.000000000e+00	3.164062500e-01	6.250000000e-02	3.906250000e-03
1.000000000e+00	7.382812500e-01	3.125000000e-01	5.078125000e-02
2.000000000e+00	9.492187500e-01	6.875000000e-01	2.617187500e-01
3.000000000e+00	9.960937500e-01	9.375000000e-01	6.835937500e-01
4.000000000e+00	1.000000000e+00	1.000000000e+00	1.000000000e+00

The probability of not detecting a violation is given below the double line for two cases, p=0.5 and p=0.75. This represents the chance of making a wrong decision. When p=0.5, the error rate is 68% and when p=0.75 is about 26%. Is this a tolerable amount of error? Perhaps for p=0.75 but not for p=0.5. Thus 4 samples would not be adequate for p=0.5.

Rather than trying to specify a single number as small, it is better to consider what background levels are likely to be (what proportion are in violation) then decide what magnitude of violation is detrimental to the environment. From this sample size can be chose to make error rates small.

2. Criteria for Defining Nutrient Over-enrichment

WQAAC Finding: Only few states have set nutrient standards and under the Clinton administration's Clean Water Initiative the EPA plans to issue guidelines for setting nutrient enrichment criteria. Nutrient standards are difficult to establish because the sensitivity of waters to nutrient loads varies significantly with the chemical, biological and hydrologic characteristics of the water body. The WQAAC would concur with the current criteria used by the DEQ for the short term. However, the WQAAC recommends that the Secretary of Natural Resources immediately convene a nutrients technical advisory committee (TAC) to review the EPA guidelines and to help develop criteria for defining nutrient over-enrichment. The Nutirents-TAC approach was tried in the 1980's and provided background for some of the current standards used in water quality monitoring in Virginia. Since that time, there has been a considerable amount of research and monitoring work undertaken in Virginia and elsewhere. As a result a new TAC should be able to bring much more information to bear on the questions of both quantitative and qualitative standards for the various types of waters in the Commonwealth.

In addition to the EPA activity, two factors will facilitate the work of a Nutrients- TAC. First, several states now have a decade or more experience in development and implementation of water quality standards for nutrient sensitive waters. The combined experiences provide a substantive base for evaluating Virginia's options. Second, the anticipated development of TMDLs under the Clean Water Act and the mandates of the Virginia WQMIR will move the science of estimating pollutant impacts into more standardized approaches. The use of simulation models for estimating watershed loadings, assimilation capacities, and ecological impacts is rapidly establishing a consensus for how to integrate various professional opinions into useful guidance.

The WQAAC recommends that the Secretary of Natural Resources appoint a Nutrient Enrichment Technical Advisory Committee (TAC) and task it to: (1) review other states' approaches to defining nutrient over-enrichment; (2) synthesize the current state of science with respect to defining nutrient over-enrichment; and (3) recommend specific quantitative and qualitative (narrative) criteria for defining nutrient over-enrichment in state waters. The TAC should be composed of researchers from the Commonwealth's universities, scientists from state agencies, technical representatives from selected federal agencies, and a facilitator. Background reviews of other state approaches and the state of the science would be developed by the facilitator and distributed to the TAC in advance of any meeting(s). Given an objective to achieve "best currently available guidance" there should be no need for additional research or even extended analysis. The TAC's role would be to familiarize itself with the available information and develop a consensus guidance for the Commonwealth . The entire process should be completed in one year, including review of the final recommendations.

3. Criteria for defining Shellfish Consumption Restrictions

WQAAC Finding: Regulatory responsibility for classification of shellfish growing waters resides with the Virginia Department of Health (VDH). Furthermore, these VDH regulations are enforced by FDA and compiled in a document called the National Shellfish Sanitation Program Manual of Operations. The DEQ bears the responsibility for interpreting the VDH designations as an indicator of overall water quality contamination. The VDH is necessarily risk averse in its condemnation decisions. However, VDH caution may result in relatively low priority waters for remedial action getting listed as impaired. The DEQ has considered this possibility and the new Guidelines recognize the "restricted" category as being "partially impaired".

The WQAAC views the use of the VDH categories as a risk policy decision that is not appropriate for WQAAC comment. We do find that the science on this matter is unsettled and new protocols are being developed to better assess the extent and significance of contamination. The DEQ should work with the VDH to review these new developments and to determine their applicability in Virginia and the consequences for the designation of impaired waters.

The science relating to shellfish bed contamination is still developing. Specific problems related to the validity of "classifying" shellfish waters have been debated intensely over the last 10 years. One response was the short-lived National Indicator Study. This program was expected to address two of the major issues surrounding shellfish growing areas: (1) criticisms that the current indicator system is inadequate to predict public health risk, and (2) development of improved, rapid methods for assessment of public health risk in water environments. Some progress was made but the program has lost momentum.

From a short-term perspective the widely used most probable number (MPN) methodology to measure coliforms (total or fecal) has been improved upon by developing the direct count methods involving membrane filtration. The EPA has pioneered this approach and published methods to detect fecal coliforms and *Escherichia coli.* The EPA has also proposed an alternate indicator (the enterococci) to coliforms test that is considered by the agency as "better" in marine waters and on which to base recreational and shellfish standards. The value of these methods resides not only in a significant improvement in accuracy, but also because the number of samples that would have to be collected by state agencies to characterize an area would be reduced. Studies performed in EPA laboratories suggest that one sample processed by direct membrane filtration technique requires in excess of 20 MPN processed samples to obtain the same level of confidence and accuracy for a known coliform density recovered from estuarine water. Labor costs are reduced, as well as time to obtain the final result comparing to the direct count approach. The ability to detect exceedance of standards would also be improved. Most states have not adopted these approved methods because (1) the inertia of departing from the tried and true, (2) neither the EPA nor the FDA have been especially aggressive in facilitating adoption of direct counting methods, and, (3) comparative data sets have not been developed. Virginia could lead the way by supporting the need to evaluate direct detection methods for fecal coliforms, *E. coli* and the enterococci, and by performing studies to compare direct and MPN methods to decide whether adoption of the former is warranted.

From a longer term perspective using fecal coliforms and *Escherichia coli* as indicators is a topic of considerable controversy when applied to shellfish growing areas where nonpoint sources contribute to deteriorated sanitary water quality. Because the coliform indicator is nonspecific in terms of warm blooded animal source, i. e., does not differentiate human from animal; and because it is found in soils and stormwater runoff, the interpretation of what elevated fecal coliform densities mean in terms of public health risk in many receiving waters is equivocal from a classification perspective. Thousands of acres of shellfish growing areas are thus classified and subject to closure using a system that has never been effectively validated or shown to predict public health risk. One recent approach at least to unravel part of this problem has fostered interest in methods to discriminate human from animal contamination. This approach may prove useful for identifying sources of fecal coliforms or *Escherichia coli* within a watershed. However, the identification of source still leaves open the question of choosing an indicator of public health risk necessary for water classification. Support for continuing studies in these areas is warranted.

4. Contamination of Sediments

WQAAC Finding: The role of sediment in the uptake, release and transport of pollutants, as well as sediment-bound nutrient and contaminant interactions with water and biota within the aquatic environments is well recognized. The effects of excess sediment and why it is a major concern include: 1) blocking light to the periphyton and algae, 2) smothering the macroinvertebrates and plants, 3) destroying bacterial slimes and periphyton that serve as the base of the food chain in streams, and 4) reduced dissolved oxygen levels.

Virginia, like many other states, does not have standards for sediment pollution. The toxics data that have been collected from sediment samples are compared to what are termed Effect Range-Median (ER-M) values. The ER-M value for a particular substance represents the fiftieth percentile of all sediment concentration values observed or predicted to adversely affect aquatic biota. The ER-M value represent "the concentration above which effects were frequently or always observed or predicted among most species" (NOAA technical memorandum, 1990). For 305(b) assessment, the DEQ uses ER-M values for substances that the authors of the NOAA document report moderate to high confidence. If there is no ER-M value for a substance, the ninety-ninth percentile value from the cumulative frequency distribution data for toxic substances in sediment samples collected in Virginia are used in determining use support. The ninety-ninth percentile values are used because for those substances which both ER-M and ninety-ninth percentile values are available from STORET, the ninety-ninth percentile values are closest to ER-M values. The methodology used by DEQ appears to be reasonable and should continue until more comprehensive guidelines are developed by a sediment-Technical Advisory Committee (Sediment-TAC) for sediment monitoring.

5. Geographically Identify Impaired Waters

WQAAC Finding: Monitoring stations take measures at discrete places on streams, lakes and estuaries. When a violation is determined at the location, the DEQ must determine the spatial extent of the impairment. For the short term, the current DEQ approach seems well suited to prevailing data and analytical limitations. The result is translated into impaired stream miles and the sum of the miles becomes a part of the 305(b) and 303 (d) reports. Long term improvements in current DEQ approach can be achieved by employing a statewide watershed assessment program. This would involve defining watersheds at different scales and then monitoring and modeling at those scales. The effort would be a direct contribution to the requirements of the WQMIR. The notion of stream segments would be replaced by watershed assessment.

As a first step, in a long-term strategy, the state should adopt a consistent system of defining watersheds at different (nested) scales. The logical alternative would be to employ the USGS catalogue system. With these watersheds, defined improvements in both the 303 (d) and 305(b) process could be realized.

For the 305(b) reports that are expected to describe the overall condition of the states waters, critics contend that the limited number of monitoring stations makes the calculation of percent of streams in the Commonwealth that are impaired flawed. To address this issue, the WQAAC recommends that the DEQ develop a protocol for probability based sampling of watersheds in the Commonwealth to characterize the overall conditions of the state's waters. The procedures and monitoring requirements suggested by the protocol should be implemented and the results be used for the 305(b) report.

The 305(b) process that calls for a general assessment of the states' waters and trends in that assessment may be hampered by the location of some stations. A probability based monitoring design for aggregate trends assessment would use a subset of the existing stations but would also require some new monitoring locations._For example, primary purposes for locating individual monitoring stations might include "upstream of point-source discharge," "downstream of point-source discharge," and/or "general water quality." DEQ's ability to make general statements about changes in water quality within the Commonwealth would be enhanced by efforts to gain a better understanding of how individual monitoring stations contribute to a statewide water-quality monitoring network. The agency needs to evaluate the goals of the monitoring program and the program's ability to achieve these goals. The number and locations of sampling stations should be reviewed in light of these goals. The DEQ should identify rationale for placement of individual monitoring stations, and adjust monitoring station locations (if needed) based on those rationales so as to move towards a statewide water-quality monitoring network.

The DEQ approach is to apply judgement based on watershed configuration, point source location and land cover. A possible concern is that this approach to calculating the spatial extent of the impairment around the stations is ad hoc and may be inaccurate. This inaccuracy is of most concern when developing a restoration strategy for the impaired area, because an inaccurate description of the spatial area impaired may not yield an effectively targeted water quality improvement program.

There are 2 ways possible ways to determine the extent of impairment, if the current approach is deemed unsatisfactory. However, both would bring an increase in analytical costs to the agency. The first is to extend the monitoring process incrementally until the problem is not detected. This approach might hold promise if the monitoring costs are not excessive. Significant monitoring costs might be expected if the suspected source of impairment in non point source and "event triggered" sampling is not possible. The second approach is apply models of the watershed to that are calibrated to the location and magnitude of the detected impairment. Modeling activity would, in principle, structure the assessment process by explicit incorporation of sources, land cover, hydrology, transport coefficients and the like to "back out the most likely spatial extent of the problem. Modeling that is capable of such results may need to be developed for TMDL and WQMIRA compliance, but such models are infrequently employed at this time.

5. The 303(d) and 305(b) Reports Shall Identify and Summarize Trends

WQAAC Finding: The DEQ has made significant strides in the past year in providing funds and data for the valid determination of trends and for development of software that will enable the agency to conduct trends analysis in-house. The DEQ effort is a sound response to JLARC criticism. The DEQ plans to continue this work and make the trends assessment available on the World Wide Web. The trend analysis, however, has limitations and efforts to address these limitations are needed. The WQAAC understands that DEQ is presently taking actions to address some of these issues, and the WQAAC encourages DEQ to continue these efforts. In addition, trend assessment can be made more cost effective by closer partnerships with the USGS.

Review of data during current trends analysis research (Zipper, et al.) has revealed a number of data-quality problems, in particular, values which are obviously erroneous. Problems appear to originate both from the laboratory and from data-entry procedures. The effects of such errors are several. For one thing, they cause the person seeking to utilize STORET data to question validity of other values. They also complicate the trend analysis procedure, as erroneous observations must be removed from the data set prior to analysis; in some cases, this requires a manual point-by-point analysis. Of course, erroneous data entries that are not successfully identified may affect the validity of trend-analysis procedures. Apparent data problems are typically identified because of values that appear to be quite high, relative to other values in the data set. If extreme outlying values are accurate reflections of water quality, they should be of immediate concern to DEQ. If they are not accurate, they should not be entered into STORET or other long-term data files. STORET data for water-quality monitoring stations operated by the U.S. Geological Survey appear to be free of erroneous values, such as those that characterized the DEQ data set. The DEQ should reference the data-quality assurance procedures utilized by USGS as a first-step in developing data quality-control procedures appropriate for DEQ. The DEQ should establish practices for controlling quality of data being entered into STORET.

On a day-to-day basis, values of many water-quality parameters will vary in direct response to changes of flow volumes. Statistical procedures exist for incorporating relationships to flow-volume into analyses for trend; use of such procedures will typically increase the ability to detect a trend and reduce the chance of detecting a false trend. Availability of accurate flow-volume data will then increase the probability that the DEQ will be able to detect any trends that do exist over shorter time periods. Where possible, monitoring stations should be calibrated to allow collection of quantitative flow-volume data, and these data should be collected on a routine basis during water-quality sampling visits. The USGS has made efforts to address this limitation in some of its recent reports on water quality trends in Virginia. The DEQ should form closer partnerships with the USGS in addressing this limitation of current trend assessments.

The effectiveness of trend analysis procedures depends on the precision of laboratory measurements. Information about the analytical procedures and changes in these procedures is important for evaluation of potential trends. The Division of Consolidated Laboratory Services should maintain a record of analytical procedures used during given time periods, and their precision.

Trend analyses typically utilize seasonal comparisons. Efficiency of data collection efforts will be enhanced if DEQ establishes the seasonal periods to be used for trend analysis at each monitoring station (e.g., monthly, every two months, or whatever), and scheduling sampling dates to accommodate these seasonal periods consistently. Data requirements of trend analysis procedures should be considered when establishing schedules and protocols for collecting water-quality data.

6. Fully Supporting But Threatened Waters

WQAAC Finding: Given data limitations and the cost of a false positive the new category of Fully Supporting But Threatened Waters makes sense. However, the DEQ should clarify its expected response when such a categorization is made. It seems reasonable for the DEQ to direct more attention and additional state monitoring to watersheds where there is some reason to suspect that a degradation in water quality may be underway.

The DEQ could develop a specialized funding strategy for targeting sampling and modeling of threatened waters. In this way the category of threatened leads to a specific action on the part of the state to determine the validity of concern (whether existing data are leading to a false positive finding) and to determine whether there is a continuing risk. As part of an overall cost analysis of the monitoring program the DEQ could consider establishing a flexible response capacity and funds that can be moved around the state whenever a threatened waters determination is made to provide intense evaluation over the subsequent period between 303 (d) reports.

7. Impaired Waters Due to Natural Conditions

WQAAC Finding: A key point for the DEQ is the correct assessment of a water body as being naturally impaired rather than the water body having a water quality standard that is not met because of human activity. The DEQ's Water Quality Assessment Guidance Manual for 305(b) and 303(d) Reports provides no guidance as to how a water body is to be determined as being naturally impaired. At present this is accomplished primarily by the DEQ staff using their "best professional judgment." Although this procedure has limitations, it should be continued for the short-term, but with a long-term goal of providing specific criteria by which natural impairment

25729

can be determined. In summary, identification of naturally impaired waters in 305(b) and 303(d) reports is appropriate. In the short-term, the DEQ should continue to use the best available judgment of its staff, and any other information available, to categorize water bodies as being naturally impaired. However, the DEQ should develop a strategy to provide benchmark data on the natural range of water quality parameters for these water bodies. That effort would allow an accurate, justifiable and quantitative identification of naturally impaired waters and also would indicate when water quality is below expected natural conditions.

The Environmental Protection Agency and SB 1122 directed the DEQ to add a category to its 305(b) and 303(d) reports that includes waters identified as being "naturally impaired." This designation is for those waters that are assessed as being impaired, based on State water quality standards, but with the impairment being due to naturally occurring conditions that are not caused by or related to past or present human activity. Several issues are associated with this designation. The first issue focuses on the definition and correct identification of naturally impaired waters. The second issue focuses on the proper action to be taken by the DEQ once a water body is designated as naturally impaired.

A wide variety of natural impairment of water bodies may occur and must be recognized as such. Probably the most common situation is that of low-gradient streams on the Coastal Plain. It is not unusual for many of these "swamp streams" to have natural conditions that violate dissolved oxygen and pH standards. Low flow, and hence low reaeration, and high inputs of organic matter primarily from leaves from the surrounding floodplains together often lead to dissolved oxygen concentrations less than 5 mg/L, especially in the summer. The pH in the streams and other water bodies of this region also can be quite acidic (e.g. 4-6), caused by the natural input of humic and fulvic acids from the decay of leaves from the floodplains.

Other forms of natural impairment include thermal springs that do not meet water temperature standards and water bodies with naturally high levels of various chemical substances because of surrounding geochemical conditions. Natural thermal stratification also can lead to the deeper waters of lakes not meeting dissolved oxygen standards. Intense photosynthesis by phytoplankton, even when they are at natural concentrations, can cause a sufficient shift in carbonate equilibrium to cause pH standards to be exceeded. These and other forms of "natural impairment" must be recognized as the ambient conditions under which these systems function and to which their biota are adapted.

There are two aspects to the problem of identification of naturally impaired waters. The first is that the DEQ must be able to determine if a water quality standard is not being met because of natural conditions, i.e., no water quality problem exists because of human activity. This would cause the body of water to be listed as being naturally impaired, but with no further action being necessary. The second aspect is for the DEQ to be able to determine if existing water quality is below expected, natural conditions, i.e. human activity has caused a reduction in water quality below the already "naturally impaired" conditions. Further action would be required under this situation. Both of these points must be considered when developing a protocol for monitoring and assessing naturally impaired waters in order to successfully protect their natural conditions, to determine when water quality standards are not being met, and to not over-react towards excessive protection of water bodies.

A problem for the DEQ is a lack of information on the natural or reference conditions of water bodies. The range of natural conditions for water quality parameters must be known in order to identify accurately a water body as being naturally impaired. Such information would provide benchmarks against which existing conditions could be compared to determine if parameter values were within or outside of the range of natural values. The basic questions thus are "Which bodies of water have water quality parameters that naturally do not meet standards?" and "What is the expected range of values for water quality parameters in waters likely to be naturally impaired?"

To answer the above questions, the DEQ should consider undertaking an effort to provide a set of reference or benchmark data on the natural range of water quality parameters for water bodies (or types of water bodies) likely to be classified as naturally impaired. This information could be provided from existing data sets, from modeling efforts, and from special studies focused on gathering needed data on specific types of water bodies that are known to be non- or least-impacted. These benchmarks would remove the need for judgment calls in identifying naturally impaired waters and would provide the basis for justifying to other agencies and the public DEQ's classifying a water body as naturally impaired. In addition, the benchmarks would provide a means for determining if a water body actually was impaired beyond natural conditions and thus should be placed in a different category in the 305(b) and 303(d) reports.

The second issue concerning naturally impaired waters is the proper action to be taken by the DEQ once a water body is designated as naturally impaired. Inclusion of naturally impaired waters in the 305(b) and 303(d) reports, as a separate category, is justified in that this identifies those waters that do not meet standards for the reasonable and beneficial uses of those waters. The Water Quality Assessment Guidance Manual then states that no control measures, pollution reduction projects, or TMDLs should be developed for these water bodies. Any of the above actions would be wasteful of time and funds and could have the negative effect of altering natural conditions and thus being harmful to the biota inhabiting these waters. The goal should be the maintenance of natural conditions rather than the meeting of standards for uses that may not be appropriate for a given body of water.

VI. WQAAC Identified Issues

Expanded Data Bases and Analysis

WQAAC Finding: We support DEQ efforts to expand the data bases for assessments through new partnerships with volunteer monitoring groups, however the DEQ can take actions to assure the quality of these assessments. Also, we urge the DEQ to partner more closely with the USGS in preparing the assessment reports.

The declaration of a water body as impaired triggers a significant planning and regulatory program for the impaired stream segment. The expanded use of volunteer monitoring, while encouraging, requires that there be careful documentation that the data collected are consistent with the protocols employed by the agency in collecting and analyzing its own data. Until such documentation can be secured the DEQ should continue to use citizen data for placing a water body in the fully supporting but threatened category. The DEQ should provide support for training of volunteer monitors, should support studies that evaluate the quality of the data collected by citizen monitors and should support studies that will provide recommendations for expansion of citizen monitoring beyond current levels and for enhancing the quality of citizen monitoring results.

A significant gap in the assessment process is the absence of reference to the work of Federal agencies, especially the USGS. The DEQ should establish a working relation with the Richmond USGS office to share resources and data collection and assessment for preparation of the 303(d) and 305(b) reports. It is worth noting that Virginia, with the exception of the James River Basin, is completely covered by NAWQA study units (none of the studies are administered from the USGS Virginia Office). The results of the NAWQA program's water quality assessments by the USGS should be more explicitly recognized in the assessment process. Support for the USGS expansion of the NAWQA program to the James Basin should be offered to the USGS by the Secretary of Natural Resources. (See: http://wwwrvares.er.usgs.gov/nawqa/nawqamap.html)

QA/QC for Sampling

WQAAC Finding: The DEQ should provide more information on how the agency collects its "monitored" data, how frequency of sample collection is chosen, how stations are selected, and most important, are uniform quality control/quality assurance protocols in place and are they adequately enforced.

Anecdotal stories and criticisms from a variety of sources suggest the DEQ needs to be sensitive to QA/QC concerns. The only mention of this topic is found on page 3 of the Guidelines. Perhaps the state should compile/develop its own analytical procedures manual, including specific QA/QC provisions and also establish a state QA/QC office to inspect state and private laboratories which process samples from monitoring stations. The manual would be disseminated to all laboratories as well as other groups (citizen groups) who are gathering data other than "monitored."

Environmental Risk Management and Policy

WQAAC Finding: The DEQ is responsible for both environmental risk assessment and risk management. Public and professional consensus should be sought on risk assessment that applies the best available data and models. Risk management decisions are policy decisions that must rest with the administration and the General Assembly, as advised by agency professionals and stakeholders. Thus the declaration of a stream (watershed) as impaired requires a public policy decision made with reference to statistical issues, data reliability issues and the costs of being wrong in making the impairment determination. Costs of being wrong can be understood in terms of the water quality consequences of a false negative (declared an area not impaired when it is) and the budget drain on limited resources of a false positive (declaring an area impaired when it is not). The DEQ should clarify this distinction in the guidelines. A preamble that explains the assessment process in these terms would assist in communicating the results.

In addition, as the WQMIR and TMDL processes move forward the agency and political leadership, along with affected stakeholders, will need to make waste load allocation and load allocation choices and choose margins of safety. These are not technical matters although the decision should be based on the best available science.

The risk management element in water quality assessment is exemplified by the different approaches taken to inference between conventional pollutants and toxics. For toxic pollutants, instead of a statistical inference approach for small samples the DEQ uses a percentile method to determine impairments. In this case, if the 97th percentile is greater than the water quality standard, the water is listed in the 303(d) report as impaired. Letting the choice of how to use and interpret data be influenced by the consequences of being wrong reflects a policy decision. There is no "scientific" foundation to direct that choice. Similar determinations exist throughout the guidelines. As another example, the way to interpret shellfish contamination assessments in terms of whether to declare a stream impaired is a risk management (policy) decision.

The WQAAC supports development of data and models for WQMIR and TMDL plan development. However, the allocation of pollution control responsibility and the multiple implementation challenges can not be "contracted out" to consulting firms or to university scientists as an adjunct to the TMDL analysis. The water quality management process implied by the WQMIR should be carefully developed, and the role of agencies and stakeholders in making environmental policy/risk determinations clarified. The WQAAC suggests that this careful development of the planning process be part of the continuing program for water quality plan deregulation. The DEQ should request that the Secretary of Natural Resources instruct DCR and CBLAD to work directly with DEQ on this effort. The new demands of the WQMIR and TMDL process, as reflected in a revised planning process, will challenge the staff capacity and the current organizational division of responsibility for water quality management. The Secretary of Natural Resources should request a review of DEQ/DCR/CBLAD capabilities for water quality plan-implementation, and should consider organizational, managerial and resource needs for that implementation.

Determine the Costs of Current Monitoring and Assessment Activity

WQAAC Finding: As the requirements for doing assessment and making risk management decisions are increased there will be demands for increased sampling locations, increased frequency of sampling and sampling for an expanded number of pollutants. The DEQ has not completed a cost and effectiveness assessment of its monitoring and assessment efforts. The DEQ should contract for a study of the fully allocated costs of its monitoring and assessment program.

Based on a study of its existing program resources devoted to monitoring and how these resources are allocated, the DEQ should work with the WQAAC or a similar body to develop a strategic plan that will optimize the use of its monitoring funds and justify any increases needed in those funds. Such a plan would recognize statistical design issues related to station location for the 303 (d) and 305(b) process, new technologies for monitoring and opportunities for partnership with others. Specific suggestions on these topics are elsewhere in this report.

Bio-monitoring

WQAAC Finding: The use of benthic surveys by the DEQ for biological monitoring seems to be appropriate in the short-term, with the only comment being that the DEQ should look to expand the number of sites included in the monitoring program. There are, however, some important long-term issues that should be considered. They focus on a change in EPA protocols for biological monitoring and the selection of reference sites. In summary, the benthic data for DEQ's biological monitoring effort presently is collected, analyzed and interpreted in an appropriate manner. The establishment of additional monitoring sites would be useful. In the long-term, the DEQ should move toward developing the best possible reference sites and adopting the new EPA Rapid Bioassessment Protocols (RBP). Both will require additional studies within an ecoregion framework, but with the eventual outcome that the biological monitoring program will have greater power and accuracy .

The DEQ samples benthic macroinvertebrate communities for their biological monitoring of the water quality of rivers and streams. The benthic data are used to directly assess the aquatic life designated use of these waters. The ongoing program provides valuable information on water quality that monitoring of conventional pollutants cannot provide. The methods used for this monitoring program are based on the EPA's 1989 Rapid Bioassessment Protocols (RBP) for macroinvertebrates. Presently, 239 sites across the state are included in this effort, compared to 1,116 ambient monitoring sites. It would be useful for the DEQ to increase the number of sites in their biological monitoring effort, especially given the power of biological assessments compared to conventional pollutant monitoring in integrating both short- and long-term perturbations.

The DEQ's Water Quality Assessment Guidance Manual for 305(b) and 303(d) reports examines several questions concerning biological monitoring (Section 8.1). DEQ is correct to make a water quality assessment based on only one benthic survey and to rely heavily on the most recent survey. As noted by DEQ, data provided by the biological monitoring effort directly measures the response of the benthic community to cumulative perturbations and thus is not analogous to the single data points provided by conventional pollutant monitoring. A single benthic survey thus provides a strong assessment of existing water quality at a site.

The DEQ uses the criteria noted in the 1989 EPA RBP protocols to determine narrative ratings of water quality from the benthic data (i.e., not impaired, slightly impaired, moderately impaired, severely impaired). These narrative ratings are then used to categorize waters according to the ratings required for the 305(b) and 303(d) reports. The DEQ appropriately converts the RBP ratings to the 305(b) and 303(d) ratings. A rating of not impaired or slightly impaired correctly translates to fully supporting in the reports since waters found as both not impaired and slightly impaired by the RBP protocols still support a relatively healthy biotic community. Similarly, a rating of severely impaired is appropriately listed as not supporting.

The most difficult point is the handling of a water body found to be moderately impaired according to the RBP criteria. The DEQ handles this nicely by initially listing that water body as being fully supporting but threatened. The listing is changed to partially supporting if a subsequent survey continues to find the stream moderately impaired.

EPA is updating the RBP methods with new protocols. The DEQ should convert to the new protocols for the sake of consistency with other agencies and because the new protocols provide a more powerful and accurate method for assessing water quality. Basic sampling procedures will not change, but the new protocols make important changes in the way data are analyzed and interpreted. The old protocols provided specific measures (metrics) of benthic community structure to be used in all monitoring programs. Some of those metrics have been shown to be very poor measures of water quality and hence have been eliminated. In addition, EPA recognized that not all of their original metrics were useful in assessing water quality in all types of streams and rivers or all geographic regions. Thus, one set of metrics may no longer be appropriate for assessing water quality throughout the state. The new protocols provide a suite of potential metrics, but require agencies to conduct preliminary studies to determine which of the potential metrics are best at indicating water quality in a given geographic region and hence which should be used in the assessment.

An additional aspect of the required studies is the need for agencies to calibrate the newly selected metric values with the narrative water quality ratings (i.e., not impaired, slightly impaired, etc.). This is important because EPA's original correlation of metric scores with narrative ratings was subject to considerable variation among geographic regions, at times resulting in inaccurate assessments.

Both the selection of metrics to be used and their calibration will require comprehensive studies outside of the normal monitoring effort. These studies are not a trivial matter in terms of required resources. Importantly, data collected under the old protocols can be integrated with the new data into a long-term data base because sampling procedures, and hence the method of obtaining the raw data, are essentially the same for both procedures. The old data would only need to be re-analyzed using the new data analysis protocols to provide a consistent data base. By converting to the new protocols and completing the preliminary studies and metric calibrations, the DEQ will have a far stronger and more accurate method of benthic monitoring than presently exists.

Another point for consideration is that the RBP ratings of water quality are greatly influenced by the selection of reference sites. Water quality rating at a monitored site is based on the percent comparability of metric values at the monitored site compared to values at a reference site. Because of this reliance on reference sites, the choice of a reference stream will determine the impairment rating. Reference sites are supposed to reflect some "desired" water quality for the area being assessed. Because desired water quality is a policy judgment, so too is the selection of the reference site. For example, low metric scores for a reference site will result in a high percent comparability, and hence water quality assessment, for a monitored site. If the reference stream is expected to be one that is unaffected by human activity then few (if any) non-impacted rivers or streams remain in the state to serve as reference sites. In practice, reference sites will be only the "best available" or "least disturbed" sites in an area. This is acceptable if the reference stream is expected to reflect the "best attainable" water quality for the area. If the policy goal is to do better then different reference sites will need to be developed.

In the short-term, the DEQ should insure as best as possible that the reference streams used in their biological monitoring program are of the highest quality available – if that is the desired condition for the stream being assessed. For the long-term, DEQ might consider an effort whereby data from sampling of best available sites are combined in probabilistic models to construct model reference sites against which monitored sites could be compared. Such efforts are being used successfully outside of Virginia and some data currently exist that could be used to develop these model streams for portions of Virginia. Once model streams are developed, DEQ could reduce their field sampling effort at reference sites, potentially freeing resources to expand the number of higher priority monitored sites.

A final point that is pertinent to both the problem of converting to new RBP protocols as well as the need for appropriate reference sites is that of the geographic variability in the characteristics of streams and rivers across the state. The concept of ecoregions is widely used in biological monitoring, including by DEQ, to account for geographic variability among waters of different geographic areas. The waters of different ecoregions typically have different physico-chemical, hydrological and biological characteristics that can cause considerable differences among ecoregions in both the appropriate metrics to be used and in the expected reference metric scores.

Virginia encompasses six ecoregions, but with sufficient variability in some of them to require the recognition of subecoregions to adequately separate groups of similar streams. For example, the Southeastern Plains ecoregion probably should be divided into three or four subecoregions based on differences in macroinvertebrate community structure in the "best available" streams. If the problem of the need for adequate reference sites, noted above, is to be addressed by DEQ, it will need to be done so within the framework of ecoregions and/or subecoregions. Development of probabilistic model reference streams for different subecoregions would greatly enhance DEQ's biological monitoring program.

In addition, the selection of appropriate metrics and the calibration of those metrics, as required for the new RBP protocols, will need to be accomplished within an ecoregion framework. Metrics that are appropriate in one ecoregion may not be so in another. A given value of a metric in one ecoregion may not correspond to the same narrative water quality assessment in another. Integrated studies examining appropriate metrics across all ecoregions in Virginia thus is required before the new EPA protocols can be fully adopted.

Toward Enhanced Monitoring Capacity

WQAAC Finding: The Commonwealth needs to develop a strategic approach to its monitoring programs. The focus on watershed performance which now drives much of the nonpoint source management efforts, highlights the benefits of carefully siting monitoring stations so that they have maximum utility for assessing consequences of management efforts. In addition, the statistical power of various temporal sampling patterns should be carefully reviewed to design a monitoring program which will optimize analysis opportunities. The WQAAC recommends that the DEQ convene a Monitoring Technical Advisory Committee to review the monitoring program and to recommend options for modification of the sampling design in space and time. The TAC should be explicitly charged with development of a plan which: (1) ensures basic monitoring sufficient to meet regulatory program reporting needs; (2) maximizes the potential for assessing water quality management efforts; and (3) can be expanded or contracted as dictated by resources without completely compromising the data base.

The WQAAC understands that the monitoring and assessment program is pulled in many directions. We recognize that some stations will be for assessment of watershed conditions (303d) and others may be for statewide trend assessment (305b). We recognize that some waters may need more measurement attention than others. We also recognize that there are lakes as well as rivers to be monitored. The problem is how to design a sampling system that can serve these different purposes and not be unduly expensive.

For historical reasons, the design of the monitoring program reflects attention to practical matters of access to sample points as it does analytical advantages. For example, the current program is oriented to sampling related to point source discharge regulation. Today questions about non-point source discharges and general questions about overall water quality are also of interest.

Remote measurement methods

WQAAC Finding: Enhanced water quality management will require an expanded network of stations (increasing the number of sampling sites) and increased sampling frequency (more samples and increased labor). However, the implementation of this approach using the conventional monitoring methods is immediately ruled out because of the anticipated cost. There is a critical need to design and test spatially and temporally monitoring systems using the emerging technologies such as remote sensors and data transfer techniques.

The design of a cost-effective watershed-based monitoring system includes:

1. Selection of monitoring sites or stations on a watershed scale
2. Deployment of sampling instrumentation and remote sensing devices
3. Determination of sampling protocols (sampling frequency and number of samples)
4. Protocols for data transfer and database management
5. Protocols for data interpretation and analysis

In recent years, based on advances on statistical methods, and development of environmental sensors and microprocessors, there have been advances in research and technology development that address the objectives listed above. Procedures have been developed to design the monitoring network and determine sampling frequency based on statistical concepts. Research is underway to develop procedures for 'smart or dynamic sampling system', an automated sampling method that can minimize the number of samples under various hydrologic conditions without sacrificing the accuracy of estimating pollutant loads. Although sensors that facilitate continuous and remote measurement of flow rates and some critical water quality parameters such as dissolved oxygen and turbidity are being developed, much work remains to be done to refine these technologies and develop sensors that could make possible remote sensing of many other pollutants. Advances in communication and PC-based database management systems have facilitated the potential for the instant and rapid transfer, storage, analysis and interpretation, and visual display of large amounts of data.

There is a need to assemble the available information on developed technologies and then field test a remote system for cost and effectiveness. Also, a review of ongoing research and emerging technologies should be undertaken (USGS has such an activity underway at this time). The DEQ should work with relevant university experts and the USGS to assess available technologies and their application in Virginia watersheds. Such a review could be coordinated though the WQAAC. In addition, a pilot program of deployment should be undertaken by university scientists. That pilot program should be developed in partnership with private sector equipment suppliers and should lead to a complete documentation of the costs and accuracy of such systems. Experimental protocols would be developed to achieve these results.

Models for TMDL Development

WQAAC Finding: Developing a TMDL program will be required when a water body is declared impaired. Information needed to develop a TMDL program include: pollutant type, estimation of the assimilative capacity of affected water body, estimation of the pollutant loads from all sources to the waterbody, predictive analysis of the consequences of the load for the waterbody. Policy decisions that are required include determination of total allowable pollutant load consistent with water quality goals, the acceptable margin of safety in setting load targets and allocation of waste loads that will maintain the load target over time.

The TMDL implementation is expected to yield a pollutant load allocation to point source dischargers and guidelines for a nonpoint source pollution control program that could achieve the use designation goal for the water body. The TMDL process can be best developed through a watershed approach. This approach is recommended by the WQMIR and is also consistent with the Commonwealth's tributary strategies planning process. The DEQ, in cooperation with the DCR and CBLAD, needs to develop a cost-effective decision support oriented water quality modeling protocol for watershed assessments and TMDL development.

The WQAAC believes that this goal can readily be achieved by adopting tools now under development at the Commonwealth's universities. The state should develop a liaison team to work with the university community to assure that the DSS tools now under development are supported and then applied in water quality management. In addition, the WQAAC suggests that the DEQ, DCR and CBLAD establish a formal framework through which to discuss the TMDL concept in order to develop consensus on how to approach the technically difficult problem areas. There needs to be sufficient lead time in this process to enable scientists and others to develop hypotheses and seek funding for the kinds of studies that may be required. Also, an alternative approach to use of models for some watersheds – termed adaptive management - should be considered.

The EPA lists 21 models in the TMDL Guidance Report that could be used to develop a TMDL. Basically three types of models that are needed: 1) a watershed stormwater model to compute nonpoint source loads and to evaluate BMPs, 2) a water quality model for the waterbody in question, and , and 3) a cost analysis model. For the water quality module, models such as QUAL2E or WASP5 are suggested. On the other hand, the question remains which model or models are most appropriate for the nonpoint pollution computations. EPA has listed BASINS as one possible DSS overlay that can be used to integrate and employ models in an interactive way. However, use of other models developed in Virginia universities should be explored.

Modeling studies for the TMDL process can be expensive relative to management insights gained. The DEQ/DCR/CBLAD should consider an adaptive management strategy in addition to enhanced modeling for TMDL implementation. Often defensible professional judgements on stressors and solutions in some impaired watersheds can be made without the aid of models or detailed data. This is necessary when models for watershed analysis or data for their application in particular places are limited. An adaptive management strategy allows management to proceed in an uncertain environment, but requires a continuing learning process and redirection of management based on this augmented understanding. Adaptive management has been described as "learning by doing" however it is not to be construed as undirected management. Adaptive management requires explicit hypotheses about the problem to be addressed and the effectiveness of the proposed solutions. Adaptive management for water quality requires setting out actions as part of an experimental design so that the effectiveness of actions can be compared with a control area. Adaptive management requires monitoring protocols that will allow for these analyses to take place. Finally, adaptive management will require new expectations on the part of citizens and decision makers about the immediate results of any management strategy.

Water Quality Standard Setting

WQAAC Finding: There are a number of technical challenges that remain in the development of water quality standard and goals for TMDL development. The WQAAC has highlighted three of these challenges, although there may be others. The agencies of the Commonwealth should seek additional technical advice on these and other matters through a continuation of a WQAAC.

A waterbody's assimilative capacity is defined as the amount of pollutant it can hold in an assumed volume of water without exceeding the water quality standard. This "volume" has been taken generally as the volume under a 7Q10 low flow condition for rivers and estuaries. The 7Q10 has been used for decades to determine point source allocations. However, since nonpoint pollution is mainly storm-driven, EPA states that "the use of low flow-related design conditions for nonpoint source is inappropriate". Thus far EPA has not clarified the question about what is the appropriate design condition or the condition under which nonpoint sources or a combination of point and nonpoint sources should be evaluated.

Establishing TMDLs will require a higher level of understanding of ecosystem processes, especially microbial-based ones, than is obtained in current water classification and data collection/monitoring programs. Understanding ecosystem processes will involve experimental approaches, collection of sufficient empirical data, and watershed computer models for predictive purposes. Although (fecal) coliforms are a historical part of the water quality regulatory fabric, and in fact responsible for many "exceedences" of the water quality standard, they are very dissimilar from more conventional pollutants such as nutrients in at least two ways. First, the fate and effects of inorganic nutrients are generally fairly well recognized. In contrast, fecal coliforms are indirect (and imperfect) measures of the potential presence of a suite of human enteric pathogens. As allochthonous nonconservative biological elements that interact within biotic communities and also possess their own inherent physiological responses to aquatic environments, predicting their occurrence and fate is difficult and has not been extensively evaluated for a variety of natural conditions. For modeling purposes, quantitative data describing fecal coliform "dieoff" or aftergrowth coefficients will be required.. The introduction of fecal coliforms to receiving waters can vary unpredictably with season, land use, and climate. Natural populations of animals also contribute to fecal coliform loading. However, the relative risk from such pollution is unknown but considered lower than for human contamination. Predicting the fate of fecal coliforms in receiving waters is also an area that has not been studied extensively.

Appendix A: WQAAC Membership

Dr. Leonard A. Shabman (Chair)
Director and Professor
Virginia Water Resources Research Center
Virginia Tech
10 Sandy Hall (0444)
Blacksburg, VA 24061
540-231-6844
Fax: 540-231-6673
E-mail: shabman@vt.edu

Dr. Carl Hershner
Chair, Resource Management & Policy
School of Marine Science
Virginia Institute of Marine Science
P.O. Box 1346
Gloucester Point, VA 23062-1346
804-642-7387
Fax: 804-642-7182
E-mail: carl@vims.edu

Dr. Howard I. Kator
Biological Sciences
School of Marine Science
Virginia Institute of Marine Science
P.O. Box 1346
Gloucester Point, VA 23062-1346
804-684-7341
Fax: 804-642-7399
E-mail: kator@vims.edu

Dr. Eric P. Smith
Professor
Department of Statistics
Virginia Tech
410 Hutcheson Hall (0439)
Blacksburg, VA 24061
540-231-7929
Fax: 540-231-3863
E-mail: epsmith@vt.edu

Dr. Leonard A. Smock

Chairman of Biology Department
Virginia Commonwealth University
816 Park Avenue
P.O. Box 84201
Richmond, VA 23284-2012
804-828-1562
Fax: 804-828-0503
E-mail: lsmock@saturn.vcu.edu

Dr. Tamim Younos
Associate Director & Research Scientist
Virginia Water Resources Research Center
Virginia Tech
10 Sandy Hall (0444)
Blacksburg, VA 24061
540-231-8039
Fax: 540-231-6673
E-mail: tyounos@vt.edu

Dr. Shaw L. Yu
Civil Engineering & Applied Science
University of Virginia
Thornton Hall
Charlottesville, VA 22903-2442
804-924-7464
Fax: 804-982-2951
E-mail: sly@virginia.edu

Dr. Carl E. Zipper
Assistant Professor
Crop and Soil Environmental Science
Virginia Tech
364 Smyth Hall (0404)
Blacksburg, VA 24061
540-231-9782
Fax: 540-231-3431
E-mail: czip@vt.edu

25743