Final Report:
Evaluation of the Gulfwatch Monitoring Program

BY
Bruce W. Tripp, Rinehart Coastal Research Center
Woods Hole Oceanographic Institution
Woods Hole, MA 02543-1525

and

Michael Bothner, USGS Woods Hole Field Center
John Farrington, WHOI Office of Education
Anne Giblin, MBL Ecosystems Center
Judith McDowell, WHOI Biology Dept.
Peter Shelley, Conservation Law Foundation

TO
U.S. Gulf of Maine Association
Boston, MA 02116

3 October 1997

in partial fulfillment of Contract # GM 97-12

INTRODUCTION

The Gulfwatch Program incorporates some of the essential elements of a monitoring program and it is at a good point in time to reevaluate it's goals and objectives and to more sharply focus available finite resources on achievable goals. This reevaluation can be built on an extensive monitoring literature (e.g., NOAA 1991, 1997) and on the experience of other programs (e.g., NOAA 1995) In this review of Gulfwatch, we focus on a chemical contaminant monitoring program but have summarized coastal monitoring in more general terms in Appendix 2.

A comprehensive monitoring program of toxic contaminants requires understanding of:

(1) the physical processes (specifically flow characteristics) that influence the transport of contaminants;

(2) the chemical processes that influence contaminant availability, persistence, and degradation in sediments and water;

(3) the long-term biological effects that result from low-level contaminant exposure; and

(4) the human health risks associated with toxic contaminants.

The first two aspects are important in establishing realistic exposure scenarios in space and time and the last two aspects are important in linking ecological effects with concerns of contamination of resources.

In designing a monitoring program for evaluating human health risks and coastal environmental degradation, the following objectives should be considered, although not all of them are currently included in Gulfwatch:

- Define the sources of contamination (both point and non-point sources) and determine the degree to which those sources can be controlled. Assemble production and use data. This may be an extremely difficult task but if changes in contaminant inputs through changes in use patterns, new treatment technologies, or recycling efforts are to be effective, we must begin to understand the magnitude of contaminant inputs to coastal waters and how they may be reduced.
- Determine the persistence, degradation rates, and biogeochemical cycling of contaminants within coastal marine sediments and the flux of those contaminants between sediments, water, and organisms. Obviously, this requires links to current environmental research.
- Relate contaminant concentrations in the environment to ecological changes of concern.
- Describe variability of selected parameters in the natural system. Analytical, spatial and temporal variability must be described in order to interpret monitoring data.
- Improve analytical methodology and sampling design so that data of the requisite quality are generated. It is also important that a monitoring program remain flexible to incorporate the introduction of new techniques as they become available.

• A systematic and rigorous quality control-quality assurance program must be an integral component of any analytical program; including participation in inter-laboratory comparison exercises. This component is essential for the interpretation of program results.

• Data analysis and interpretation is an essential part of any monitoring program and is too often given insufficient effort. This step includes the translation of expensive monitoring measurements into coherent information that is relevant to managers.

• Public assess to monitoring results on a timely basis is important and data should be incorporated into regional and national electronic databases as soon as practicable.

Together, the above components constitute a technically sound monitoring program and the continuous incorporation of state-of-the-art scientific knowledge into the monitoring program design will ensure credible results. To avoid the trap of expending effort on simple data-collection, a well-designed monitoring program should incorporate the concept of "hypothesis testing" common to scientific experimentation. A specific testable question based on a resource management issue of concern will guide the nature and quality of measurements made. Using this approach, monitoring measurements will only be made for variables which are relevant to address the general question, "Will this measurement provide the data to permit a management decision?" This approach will create an obvious link between data collection and decision making, thus making the monitoring program more cost effective.

The kinds of testable questions to be developed fall into several categories: contaminant characterization, input quantification, near-field, short-term effects (e.g., respiratory effects) and far-field, long-term effects (e.g., community alterations). Once testable questions are developed, they are organized into a "tiered" monitoring strategy (e.g., Zeller and Wastler, 1986; NRC, 1990). A similar approach based on measured effects has also been described by Phelps et al (1987).

This monitoring strategy will generate only that information which is needed for decision-making and will not resolve other questions that may arise. It is important to avoid the natural tendency to overuse monitoring measurements by applying them to questions that were not asked when the data were gathered. A well-designed monitoring program should provide a framework within which resource management agencies can address specific management issues. The tiered approach to sampling and analysis encourages a conscious decision at each step (tier) concerning stepping up to the next level of resolution (and expense). At each point, the monitoring staff must decide if temporal and spatial resolution is adequate to address the question asked and if data resolution is of adequate precision for a management action. This tiered approach will also discourage the automatic use of expensive high-tech measurements until they are clearly needed. Scientific uncertainty should be explicitly considered at each tier level when considering adequacy of data in hand.

The tiered approach described is organized into a monitoring plan to be used by the agencies now making less organized measurements of coastal contamination. In one sense, this monitoring plan will always be evolving because we do not yet fully understand the natural processes (physical, chemical, and biological) that control the transport, fate and effects of contaminants in coastal systems. Through the adoption of this approach, it will become clear to

the monitoring program staff why each tier is a necessary base on which to build the next tier of data of increased resolution (and cost). Initial tiers represent observational data and simple measurements of bulk parameters. Depending on the response provided by such data to the testable question posed, a decision will be made whether or not to continue to the next tier. Proceeding from tier to tier will generate data of increasing resolution, accuracy and cost. Thus, more sophisticated data will not be sought unless a specific need is identified at the preceding tier level. Ideally, the more detailed studies required in the later tiers will be carried out only when simpler studies conducted in an earlier tier reveal that data of higher resolution or reduced uncertainty are necessary. Continual close links to ongoing academic research are essential so that the monitoring plan can be modified as new information becomes available. One possible mechanism for creating such links could be the application of graduate thesis research to Gulf monitoring questions that cannot be directly addressed within the monitoring activities.

GULFWATCH REVIEW

GENERAL

In general the review panel was impressed with the effort Program staff have made to date, especially given the logistical problems of running an international program with very limited funding. Gulfwatch reports have honestly detailed problems encountered and they have largely done a good job with corrective actions to address these problems. This review details a number of comments and concems both about the underlining assumptions made and about the program implementation but these critical comments are intended to guide constructive program modifications and should not be interpreted as a negative opinion of the entire program.

An underlying "flaw" in the Gulfwatch Monitoring Program is that it has been seriously underfunded from the beginning. Some of the criticisms contained in this review relate directly to that reality. The Initial Plan developed in 1991 contains many of the appropriate program elements and recommends a realistically expected level of effort for a regional monitoring program. However, only a small portion of that Plan was funded. In addition, the original Plan was overly dependent on splicing together existing local programs into a regional entity without a critical review of those programs. While this may have been a logical reaction in the face of known fiscal constraints, a collection of local programs that meet local needs cannot automatically be knitted into a regional program. Also, the myriad of differing objectives and variable data quality found in a collection of local programs make the integration effort extremely difficult. Objectives of the 1991 Plan are overstated and create a level of expectation that cannot be fulfilled. The 1991 Plan ignores the essential integration step which would continually review data produced in the context of questions asked and modify the program in the light of this review. The publication of data files is but one part of a monitoring program and this missing integration step has cascaded from the Plan to program implementation.

The formal review represented by this report needs to be incorporated as an integral element of Gulfwatch. In the initial five years of operation, a reasonable data baseline has been produced and this serves as a foundation on which to build a stronger program. The Review Panel hopes that this review will catalyze efforts to reestablish the underlying purpose for

monitoring in the Gulf, and to restate the management questions which supposedly drive the monitoring. Implicit in all of this is the continued support of a Gulf-wide regional monitoring program, including necessary resources.

FUTURE PROGRAM DESIGN MODIFICATIONS

Hypothesis Testing Now that direct field experience has been acquired and the strengths and limitations of this approach are specifically known, it is appropriate to return to ask a specific testable question. This is a critical step, from which the entire program should develop. There are differences between monitoring to measure compliance with existing requirements (e.g., to meet outfall permit requirements) and monitoring to determine spatial and temporal trends in contaminant distributions. Both are important activities but data collected for one purpose are usually not applicable for another and the goals for data collection must be clearly stated at the outset. The incentive to overuse monitoring data will be reduced if both data collectors and resource managers clearly understand the inherent limits on any data set. These limits become apparent if clear statements of purpose are included with the data.

During this just-completed initial implementation phase of Gulfwatch, a necessary focus on the mechanics of starting a new program has somewhat obscured this very basic issue. The Gulfwatch program has responsibly evolved in response to realities of the field sampling but this response has focused on the logistical and programmatic details while losing sight of the higher level mission questions. The program has reacted to logistical issues and made methodological adjustments well. It is now an appropriate point in time to use the 5-year experience base to reassess program mission and as well as methods.

Sampling in Environmentally Relevant Scales As Gulfwatch evolves, and if it expands in scope beyond tissue sampling, samples need to be taken with an understanding of ecosystem complexities and the processes governing variability in each compartment of the ecosystem if meaningful data are to be generated. For example:

Water : The coastal water mass, driven by storm and tidal energy, is changing vertically and horizontally on the time scales of minutes-to-hours; extremely low concentrations of toxic organics in sea water cannot be detected easily by standard analytical methods.

Biota: Tissue concentrations can vary with physiological state, age and health of the organism and season. The organism's location in the food web and habitat characteristics can have an effect on contaminant distribution and concentration.

Sediment : Concentrations of organic contaminants in sediments are usually correlated with the organic carbon content and grain size distribution; therefore a sampling strategy for organics should first assess such bulk parameter distribution. Muddy, high carbon sediments are usually a long-term reservoir of toxic contaminants and should be the primary sampling focus. For nutrient-related issues, sediments are the site of major nutrient transformation within the system and more frequent sampling may be required to resolve the rates and extent of nutrient biogeochemistry. A sampling strategy involving sediments and may require sampling in different spatial and temporal scales to resolve issues of toxics, nutrients and pathogens.

One specific Gulfwatch issue related to sampling with respect to environmental complexities relates to clearing of the gut prior to analysis. Evidently, the program decision not to allow the guts to clear was to provide parity with NS&T as well as to simplify the procedure. This decision presents a problem in areas with high suspended sediment loads. The observed order of magnitude difference in aluminum (Al) concentrations clearly represents something other than body burden but this question cannot be resolved with existing data. Perhaps this is an area where further testing could be done; one possibility is to routinely allow the caged mussels to clear their guts before analysis (if this aspect of Gulfwatch continues; see discussion below) and then test deploy them in areas where high suspended loads are suspected to be causing a problem. If this alternative is attempted, possible loss of organics with gut clearing must be considered. As collected, the caged mussel data are not directly comparable to native mussels for some questions, in any case. This gut-clearing issue is also related to data that was omitted from analysis because the iron (Fe) and aluminum concentrations were unexpectedly high. The logical explanation offered was that gut contents of sediment were probably high. Perhaps the tissue data could be salvaged by analysis of the Metal/Al or Metal/Fe ratios? It would also help the interpretation of metal concentrations to know the concentrations in bottom and suspended sediments at the sampling site. A reassessment of the program should include reconsideration of sediment sampling in some cases.

Gulfwatch and NOAA National Status and Trends (NS&T) sampling periods are offset by 6 months. Previously published studies have shown that seasonal patterns exist in mussel metal body burdens and concentrations (Phillips, 1976, 1980). Mucklow (1996) reports a similar result in the Gulf. For most metals, Gulf concentrations were higher in the spring than in fall. In addition, differences in the physiological state of the mussels in spring vs. fall may affect contaminant burden and there are also likely to be seasonal differences in inputs (although the NS site studied by Mucklow was located near a sewage outfall which may have been a relatively constant local source of metals). Mussels removed some distance from local sources of pollutant may reflect contaminants from distant sources in the spring when river discharges are at their highest and therefore show an even greater seasonal change in concentration. These environmental and physiological factors make the direct comparison of data between NS&T and Gulfwatch difficult. In addition, there appear to be some differences in the analytical methods which are discussed below. One suggestion might be for Gulfwatch to sample the NS&T sites more often in one year in order to provide information on seasonal variation at these sites and allow for a direct comparison of data with the labs running the NS&T samples. Gulfwatch certainly should regularly participate in the inter laboratory comparison exercises conducted by NOAA.

Benchmark stations were chosen in an effort to quantify interannual variability. It would seem best to choose these stations to represent fairly clean sites which respond to regional scale pollution inputs. There may also be strong reasons to choose benchmark stations in heavily contaminated areas to address specific management questions; for example, Deer Island would have value as a benchmark for many questions relevant to the Gulf. By monitoring these benchmark stations annually there should be a reasonable chance of picking up regional trends in 3-5 year time scales. However, as presently designed, the number of benchmark stations is fairly small and it's not obvious that they are sufficiently removed from local contamination sources that

might skew their value as representative of the Gulf-region. It would be good to consider moving some of these stations and increasing the number if this variability detection objective is to be met.

Reviewers note that the reported limit of detection for most analytes is higher than that of NS&T. Trace concentrations in samples require more elaborate, often non-standard analytical methods that account for analytical artifacts and recovery levels (e.g., high resolution, low detection limit, negligible "blank" concentrations, etc.). This issue must be resolved by Gulfwatch through improved QA/QC practices and inter laboratory comparison before data can be produced in relatively clean areas to resolve the interannual question. A regional program such as Gulfwatch has an opportunity to focus on regional issues such as long-distance atmospheric transport that local programs, with their inherently narrower focus, normally cannot. Some contaminants have a distant source (Pb from gasoline, PCBs, etc.) as well as local sources and by selecting an adequate number of stations, these multiple sources might begin to be sorted out. Such questions are prime candidates for partnerships between a monitoring program and academic research efforts.

Overall, the issue of benchmark stations should be re-examined by Gulfwatch. By only monitoring the other sites every 3 years it will take a decade or more to determine if an individual site is getting cleaner or more polluted. This seems to be an acceptable time-frame, given the regional nature of the program and the current funding level, but it also implies a continued long-term source of program support.

Links to On-going Research The bivalve monitoring concept has been successfully tested at local, regional, national and international scales for more than three decades. The best of these programs have allocated adequate resources for communication of results and for QA/QC activities. These successful programs have also kept up-to-date with these activities by maintaining close links with academic researchers who are also working in the area of coastal environmental quality. Successful monitoring programs are designed to make as simple a measurement as possible, but monitoring measurements must be interpreted in the context of ecosystem complexities. Data interpretation and the decision of when to move up to more expensive measurements would benefit close dialog with active research projects. The search for simple answers to complex questions is a recipe for failure. Newer methodologies and techniques may be necessary to address specific questions and monitoring program links to academic research is essential if the program is to make use of (or even be aware of) these new methods. Interesting peripheral questions and some of the "high-tech" analyses can be addressed through partnerships between the monitoring program and academia.

Gulfwatch is ideally situated to build these connections but does not seem to have done so to date because of the initial phase focus on start-up details. These links need to receive much higher priority as the program matures and will require program resources.

SAMPLING and ANALYSIS

Appropriate Sample Collection and Sample Handling The choice of methods for measuring nutrient or toxic contaminants and for enumerating pathogens will be dependent on the

nature of the management question being addressed (e.g., regulatory requirements or a more detailed analysis of sources and fates). Monitoring staff should be prepared to choose appropriate methodologies and to conduct a quality control/quality assurance program for all measurements made. In many cases, a lower resolution "standard method" may be required by regulation. The high resolution method may be used to provide higher precision and accuracy in circumstances where that level of analysis is warranted to address a specific question. These analytical methods are always more expensive and often cannot be conducted using the same apparatus or even the same laboratory space. A less sensitive analytical method may be adequate for a baseline survey and screening techniques that measures some bulk parameter may be a sufficient analytical tool in many cases. The choice of appropriate analytical methodology should be made in reference to already-defined management needs and this step needs to be reevaluated by Gulfwatch.

Ancillary Measurements The use of "transplanted" mussels in cages can be a powerful adjunct to sampling native populations to address some scientific or management questions. Caged mussels have been used by Gulfwatch to compare native tissue burdens to uptake by transplanted animals. Although the 5-year report discusses in great detail possible reasons why the transplanted mussels may have overshot the native mussel concentrations, the most likely conclusion that these results are due to an experimental artifact still needs to be ruled out. The caged mussel aspect of the program is time consuming and expensive. The data obtained from the transplant experiment does not seem to be based on any specific management question. The use of caged mussels by Gulfwatch should be reconsidered and incorporated into the program only when uptake data is required for a previously identified purpose.

It is appropriate that growth and Condition Index (CI) be monitored, but there is little hope that these measurements will ever be linked back to contaminate burdens except in grossly polluted sites. CI data reported by Gulfwatch is very variable and does not permit unequivocal conclusions. Growth is strongly affected by food availability and other factors that are beyond the scope of this monitoring program. The use of newer technologies (e.g., cellular or molecular indicator of exposure) seems to be a better approach, and the introduction of such new techniques to Gulfwatch will require strong collaborative links between the program and academic researchers. One specific question arises related to growth that should be resolved by Gulfwatch before drawing conclusions from the data: mussels of similar size are being collected but do we know they are also of similar age? This detail would be useful for interpreting the body burden data. Growth data has been produced but it is not discussed in the 5-year report; even if these results are confusing, the attempt deserves discussion (even if only as a reassessment of this parameter as an included component of the monitoring program).

Necessary and Sufficient QA/QC QA/QC and inter laboratory comparison results are an essential component of monitoring reports and must be integral to the conclusions drawn. Quality control protocols are well-described in Gulfwatch reports for the metal and organic analyses but the information on standard reference materials, limits of detection, blank analyses and spike recoveries are missing. What is the comparison between the observed concentrations of metals or organics and published (accepted) concentrations on standard reference biological materials, for example? Apparently, Gulfwatch is generally following the analytical procedures of NOAA NS&T and has participated in some inter-laboratory comparison exercises. However, the results

of this work are not reported in the Gulfwatch materials reviewed. The fact that appropriate QA/QC has apparently been carried out is to be applauded, but reporting these results is an essential and integral part of the tissue concentration interpretation. Gulfwatch data at this point are uninterpretable (and therefore unreviewable) because this essential information is missing. Articles by Taylor, (1985, 1985a) provide good overviews of this critical topic. Community standards exist for this presentation (NOAA, 1993; UNEP, 1990, Villeneuve and Mee, 1991 and 1992). In addition, NOAA and UNESCO have published manuals for the use of standards and reference materials (NOAA, 1994; UNESCO, 1990; UNEP, 1990; UNEP, 1991).

Reviewers identified a detection limit difference between NOAA and Gulfwatch above. Gulfwatch partially addresses this detection limit issue in a discussion the 5-year report. This summer, we supervised a student intern who attempted to use a combination of Gulfwatch and NOAA NS&T data to create a combined database in an effort to elucidate regional trends in space and (if possible) time. We found that the limits of detection in the Gulfwatch data are higher than those of NS&T, making it impossible to create a usable combined dataset except in highly contaminated areas such as Boston Harbor. The incomplete QA/QC discussion in Gulfwatch reports does not permit a user to identify the precision or accuracy of reported concentrations, rendering the dataset useless for many purposes. This important issue deserves vigorous attention by Gulfwatch and the application of a significant portion of Gulfwatch resources to an on-going QA/QC effort.

DATA INTERPRETATION

Data interpretation One underlying objective of any monitoring program is to make interpreted information (as opposed to raw data) understandably available to those who need it for decision-making. The production of a data table of analytical results is a small part of the monitoring program. To assure that data are properly used for management decisions, it must be interpreted in the context of the management question asked, taking into consideration relevant environmental complexities. Interpretation of contaminant data will require synoptic information on physical, chemical and biological processes that affect the degradation and transport of these contaminants. Information such as currents, tidal stage, temperature, salinity, light penetration as well as other parameters will be necessary to place contaminant concentration data in an environmental context. Given existing Gulfwatch spatial and temporal sampling schedules, unknown natural variability and present analytical skills, trend detection should only be expected where order-of-magnitude differences are observed. In the recommended program reassessment, Gulfwatch can certainly increase the level of detail of any of these issues but this improvement will come at a cost and should be undertaken only for specific reason.

Data Storage and Security This important subject is not addressed in Gulfwatch reports. Data storage, on multiple media and in more than one location, is essential for long-term security. Sample archives (tissue and extracts) are also "data" and their secure storage should be seriously considered (and the archiving expense seriously addressed). Is there currently a provision for archiving samples for the long-term? Old samples could be invaluable to assess the magnitude and history of changes in contaminants that are not detectable with present analytical tools or not measured because the linkage with adverse biological effects have not yet been identified. Old

samples also represent point-in-time analytical opportunities that cannot be repeated. This issue is admittedly beyond the scope of the present Gulfwatch program but should be incorporated into their long-range planning.

Interpretation in Context of Management Question Asked Data collected in a monitoring program ideally focuses on a specific management question asked. Research programs may complement and supplement the program database, but the focus of program interpretation should remain on clearly stated management issues. As discussed above, the interpretation of monitoring data must include awareness of natural system complexities, natural variability and scientific uncertainty. Gulfwatch efforts have not maintained this focus as rigorously as it could have. One underlying objective of a monitoring program is to make information (as opposed to raw data) understandably available to those who require it for decision-making. One mechanism to maintain this focus is to give a high priority (and sufficient program resources) to data visualization. Spreadsheets of raw data are not normally useful to resource managers and significant program effort should be directed toward this goal. At this point in time, Gulfwatch is playing "catch-up" with the release and interpretation of monitoring data. The draft 5-year retrospective report (Jones et al, in prep) is the major vehicle to accomplish this and should be given top priority (discussion of this 5-year report follows below).

This particular criticism of data interpretation needs to be placed in perspective because the program has achieved goals that also are not adequately highlighted in reports reviewed. Two "management question" are: "What are the region-scale trends in chemical contaminants over a decadal time frame?" and "Is there a significant change in environmental concentration of selected chemical contaminants?" Gulfwatch has produced a baseline of contaminant distribution information that is essential to address these management questions. Due credit must be given for this result and review panel recommendations made to modify the program should not obscure this basic success.

RECOMMENDATIONS

The questions addressed in the Gulfwatch annual reports largely focus on testing hypotheses which relate to the methodology, i.e. seasonal variation in contaminant burdens and tissue concentration similarity between caged and native mussels. These questions were appropriate in the early phases of the program but it is appropriate at this time to reassess the underlying rationale for monitoring in the Gulf of Maine and to move on to broader questions. The review panel expects that some program modification would result from this assessment.

Gulfwatch will probably not ever be supported at a level where currently identified and desired monitoring tasks can be carried out by this program. With this as a given, then efficient use of data from other sources must be incorporated into this effort. One important management issue is how do contaminant burdens in the mussels relate to inputs. Obviously, direct measurement of inputs by any program funded at the level of Gulfwatch is out of the question, but it may be possible to take advantage of programs where inputs are being measured on a regular basis, such as Boston Harbor. By following year to year variations in mussels near a monitored source of inputs, insight in the relationship between loading changes and body burdens might be

better understood. Gulfwatch can make use of estimates of inputs derived from other sources and incorporate those estimates into the program. Use of non-program data is a necessary activity when program resources are extremely limited.

Data being generated by Gulfwatch has value and that value can be used to leverage partnerships with regional (state and provincial) resource management agencies and with academic research programs. Data from Gulfwatch need to be interpreted in the context of mussel watch data from other programs and to do this, Gulfwatch QA/QC issues need to be resolved. There is feeling from the printed Gulfwatch reports that the program has leveraged only limited resources from outside the principle EPA grant but Gulfwatch now has the experiential base to be in position to do so. Incorporation of non-Gulfwatch information into Gulfwatch data interpretation should also be pursued.

As a part of this review exercise, the panel has reviewed a preliminary draft of the 5-year Gulfwatch retrospective report (Jones et al, in prep) and makes some general comments on that draft here. When this report is revised, we urge that it be distributed for a critical peer review. This report summarizes the start-up phase of Gulfwatch and deserves scrutiny and comment by a group of experts in the field.

Comments made in the above discussion are relevant to the reworking of this draft report. In addition, the draft reviewed by the panel could benefit from a reorganization that explicitly addresses a series of questions:

What did Gulfwatch intend to accomplish?

What did we find?

What questions can we answer with data collected? (equally, what questions cannot be answered?)

How can we improve the program?

From this reorganization should come a sharper focus on a review of Gulfwatch activities and achievements and a reduction in discussion of more general topics.

In addition to this full report of the initial five years, Gulfwatch should also produce a short summary version written for a general audience that explains program objectives and relevance, thereby justifying the mussel watch approach in this region and demonstrating that program goals are being met. This summary is a critical publication that should highlight the value of the Gulfwatch program to the larger user community and be targeted to managers, legislators, fishermen, the aquaculture industry, educators, and the interested public. The program needs to get the information, in condensed non-scientific terms, into the hands and plans of those who benefit from knowledge of the health of the coastal environment. They will become the critical constituency for justifying and maintaining the program.

The draft 5-year retrospective report is the most important document produced by Gulfwatch and it needs to clearly state, without overstating, the results and conclusions of the program. Several specific points to consider in revision of the draft 5-year report:

• the program has necessarily focused on techniques and methodology in this start-up phase but now it will reassess it's objectives and the means to meet them. It will incorporate

an on-going self-evaluation process as an integral program component. Gulfwatch has reasonably evolved to meet the realities of field conditions and now needs to apply that same flexibility to refocus more sharply on management issues. It is time to restate the original premise of Gulfwatch, based on the initial five years experience.

• data reporting, including data interpretation, has occurred in a burst at the end of this 5-year period and needs to be given equal priority with field sampling and sample analysis. Data reporting necessarily includes data synthesis (and "translation" for non-scientists). These components need to be an integral part of the program; presently there is a serious gap between analytical results and information transfer and to date there has been no process developed to address this problem. The reviewers note that this is a problem inherent in the original Plan that is only reflected in the initial implementation effort.

• focus on the regional questions and the regional scale of the dataset without over-justifying the program or over-interpreting the data. Over-interpretation could ultimately put the entire program in jeopardy. At all costs, avoid using the retrospective report as a data dump which attempts to summarize all knowledge; maintain a sharp focus on this region and these data. After five years, Gulfwatch has produced a reasonable baseline of contaminant concentrations. This baseline is an acceptable result of the effort made; it should be highlighted as the product of this program and does not need to be oversold. Gulfwatch data does need to be compared to data from other sources (e.g., FDA market basket data) to get a better idea of how it can be used and to see if Gulfwatch really is addressing the overall management issues. For example, what are the levels of contaminants in other harvestable species from the Gulf?

• review and critique the original Plan objectives, placing the Gulfwatch effort in the perspective of the full Plan. Plan objectives are overstated and omit critical integration steps that are criticized by this review panel; the 5-year report should address these plan deficiencies and highlight accomplishments that have been achieved.

• discuss how to improve QA/QC efforts, including reporting of QA/QC results. Issues such as limit of detection of organics and analytical methods used for Hg and Pb require investigation and should incorporate collaboration with analytical chemists who are at the forefront of these analyses.

• Two issues are highlighted here as examples of how all monitoring issues are linked and how Gulfwatch (including program reporting) could be more sharply focused on the region. The question of mercury (Hg) is one example. The review panel concludes that comments on Hg in the draft 5-year report are suspect until QA/QC issues are addressed. Hg is a difficult analysis and the reported standard error makes the Gulfwatch dataset questionable. What analytical method was used? Were "standard sediments" analyzed? Was Hg included in the inter laboratory comparison exercises? Until such questions are answered, caution should be used in reaching conclusions concerning trends in this metal. Additionally, how does this data fit into the national discussions related to possibly lowering the current public health standard for Hg? If there is no particular management issue at stake, does the trend of Hg matter?

In a similar vein, we have argued in this review that a regional monitoring program is in a unique position to address regional issues and suggest that this role be given greater emphasis in any future reassessment of Gulfwatch. Reviewers have applauded the links Gulfwatch has made to an existing national monitoring program where such links provide appropriate structure and oversight but Gulfwatch should not lose sight of the fact that requirements of a national program may give less than adequate weight to regional issues. For

example, it may be appropriate from a national perspective to place little emphasis on specific contaminants that could be of importance in the Gulf region. NOAA S&T has opted not to measure some of the less common high molecular alkyl PAHs that are found in quantity in some crude oils. If these particular oils are imported into the Gulf region for transmission via the Portland pipeline, then the measurement of such chemicals would be appropriate in this regional program.

Any program reassessment that results from this review should look beyond the immediate questions of mussel watch implementation to broader issues such as those provided here.

CONCLUSIONS

The "mussel watch" approach to coastal monitoring is valid in the Gulf of Maine region and in general it is being competently and responsibly applied by Gulfwatch. The start-up phase of Gulfwatch has focused heavily on the mechanical details of implementing a field program at the expense of management questions and the program is at an appropriate juncture to reestablish underlying program objectives. To accomplish this, program staff should broaden their network of active advisors to include research scientists and practicing resource managers.

In general, Gulfwatch objectives match those of the 1991 Plan, but the Plan is very generic and contains overstated goals that should be reexamined. The funded Gulfwatch program is a remnant of the monitoring program recommended in the Plan and some of the deficiencies identified by this review relate directly to the original plan and to a lack of sufficient funding to implement a full monitoring program.

The present Gulfwatch Program is but a part of a necessary coastal monitoring effort in the Gulf of Maine region. The current program activities and the present mix of funding support is a legitimate initial effort but should not define coastal monitoring in the Gulf of Maine. Program staff should not focus exclusively on present "mussel watch" activities or on existing identified support resources but should reach out to the wider scientific and management communities to develop a monitoring program that meets the needs of those communities and is supported at a level consistent with the value of the natural resources being protected. Gulfwatch should consider the incorporation of other tools as needed, perhaps in concert with academic research projects.

It is essential that a regional monitoring effort be continued over the long term and that this activity be reviewed regularly by outside experts to assist program staff with their mission.

ACKNOWLEDGMENTS

This review has been supported by a grant from the Gulf of Maine Council on the Marine Environment and the U.S. Environmental Protection Agency with supplemental support from the WHOI Rinehart Coastal Research Center.

REFERENCES

CAPUZZO, J.M. AND KESTER, D.R., 1986. Biological effects of waste disposal: Experiential results and predictive assessments. In: Ocean Processes in Marine Pollution, Vol. 1, J. Capuzzo and D. Kester (eds.). Krieger Publ. Co., Malabar, FL.

CAPUZZO, J.M., MCELROY, A.E. AND WALLACE, G., 1987. Fish and shellfish contamination in New England waters: An evaluation and review of available data on the distribution of chemical contaminants. Coast Alliance Special Report, Boston, MA.

CAPUZZO, J.M., MOORE, M.N. AND WIDDOWS, J.W., 1988. Effects of toxic chemicals in the marine environment: Predictions of impacts from laboratory studies. Aquatic Toxicol. 11:303-311.

FARRINGTON, J.W., GOLDBERG, E.D., RISEBROUGH, R.W., MARTIN, J.H. AND BOWEN, V.T. 1983. U.S. "Musselwatch" 1976-1978: An overview of the trace metal, DDE, PCB, Hydrocarbon and artificial radio nuclide data. Environ. Sci. Technol. 17: 490-496.

FARRINGTON, J.W. AND WESTALL, 1986. Organic chemical pollutants in the oceans and groundwater: A review of fundamental properties and bio-geochemistry. In: The Role of Oceans as a Waste disposal Option, G. Kullenberg (ed.), NATO ASI Series No. 172, Proceedings of a Workshop in Vilamoura, Portugal, April 1985. D. Reidel Publ., Dordrecht, Holland.

FARRINGTON, J.W., DAVIS, A.C., TRIPP, B.W., PHELPS, D.K. AND GALLOWAY, W.B., 1987. Mussel Watch: Measurements of chemical pollutants in bivalves as one indicator of coastal environmental quality. In: New Approaches to Monitoring Aquatic Ecosystems, ASTM STP 940, T.P. Boyle (ed.), Philadelphia, PA.

JONES, S.H., M. CHASE, J. SOWLES, W. ROBINSON, P. HENNIGAR, G. HARDING, D. TAYLOR, P. WELLS, J. PEDERSON, K. COOMBS, K. FREEMAN, L. MUCKLOW. in prep. The First Five Years of Gulfwatch, 1991-1995: a review of the program and results.

MUCKLOW, L.C. 1996. Effects of Season and Species on Physiological Condition and Contaminant Burdens in Mussels (Mytilus edulis and M. trossulus): implications for the Gulfwatch Program. Tech. Rept., Gulf of Maine Council on the Marine Environment.

NOAA. 1984. National Status and Trends Program, A Progress Report and Preliminary Assessment of the Benthic Surveillance Project. NOAA-NOS, Rockville, MD.

NOAA. 1987. A Summary of Selected Data on Chemical Contaminants in Tissues Collected During 1984, 1985 and 1986. NOAA Tech. Memo. NOS OMA 38, Rockville, MD, USA.

NOAA. 1989. A Summary of Data on Tissue Contamination from the First Three Years (1986-1988) of the Mussel Watch Project. NOAA Tech. Memo. NOS OMA 49, Rockville, MD, USA.

NOAA. 1991. Mussel Watch Worldwide Literature Survey - 1991. Ed. A.Y. Cantillo. NOAA Tech. Memo. NOS ORCA 63, Rockville, MD, USA.

NOAA. 1991a. Second Summary of Data on Chemical Contaminants in Sediments from the National Status and Trends Program. NOAA Tech. Memo. NOS OMA 59, Rockville, MD, USA.

NOAA. 1993. Sampling and Analytical Methods of the National Status and Trends Program, National Benthic Surveillance and Mussel Watch Projects. Vols. 1-4. NOAA Tech. Memo. NOS ORCA 71, Rockville, MD, USA.

NOAA. 1994. Use of Standards and Reference Materials in the Measurement of Chlorinated Hydrocarbon Residues. NOAA Tech. Memo. NOS OMA 77, Rockville, MD, USA.

NOAA. 1995. International Mussel Watch: initial implementation phase final report. NOAA Tech. Memo. NOS OMA 95, Silver Spring, MD, USA.

NOAA. 1997. World Mussel Watch Database. NOAA Tech. Memo. NOS ORCA 109, Silver Spring, MD, USA.
NRC. 1990. Managing Troubled Waters. National Research Council, Publications Office, National Academy of Science, Washington, D.C.

OFFICE OF TECHNOLOLGY ASSESSMENT. 1987. Wastes in the Marine Environment. OTA-0-334, Washington, DC.
PHELPS, D.K., KATZ, C.H., SCOTT, K.J. AND REYNOLDS, B.H. 1987. Coastal monitoring: evaluation of monitoring methods in Narragansett Bay, Long Island Sound and New York Bight, and a general monitoring strategy. In: New Approaches to Monitoring Aquatic Ecosystems, ASTM STP 940, T.P. Boyle (ed.), Philadelphia, PA.
PHILLIPS, D.J.H. 1976. The common mussel, Mytilus edulis as an indicator of pollution of zinc, cadmium, lead and copper: I Effects of environmental variables on the uptake of metals. Mar. Ecol. 22(2):187-198.
PHILLIPS, D.J.H. 1980. Quantitative Aquatic Biological Indicators. Applied Science Publishers, London, 488 pp.
TAYLOR, J.K. 1985. Standard Reference Materials: Handbook for SRM Users. National Bureau of Standards Special Publication No. 260-100. U.S. Dept. of Commerce.
TAYLOR, J.K. 1985a. Principles of Quality Assurance of Chemical Measurements. National Bureau of Standards Tech. Rept. NBSIR 85-3105. Gaithersburg, MD.
UNEP. 1990. Contaminant Monitoring Programmes Using Marine Organisms: quality assurance and good laboratory practice. Mar. Pollut. Studies No. 57.
UNEP. 1990. Reference Materials and Methods: a programme of support for regional and global marine pollution assessments.
UNEP. 1991. Sampling of Selected Marine Organisms and Sample Preperation for the Analysis of ChlorinatedHydrocarbons. Ref. Methods No. 12/Rev.2.
UNESCO 1990. Standard and Reference Materials for Marine Science. Manuals & Guides # 21.
Villeneuve, J.-P and L.D. Mee. 1989. Chlorinated Hydrocarbons in Tuna Homogenate IAEA-351: results of a world-wide exercise. ILMR Intercal. Rept. No. 44. Monaco.
Villeneuve, J.-P and L.D. Mee. 1992. World-wide and Regional Intercomparison for the Determination of Organochlorine Compounds and Petroleum Compounds in Sediment IAEA Sample 357. IAEA Marine Environment Laboratory Report No. 51. Monaco.
ZELLER, R.W. AND T.A. WASTLER, 1986. Tiered ocean disposal monitoring will minimize data requirements. In: Oceans 86, Vol. 3 Monitoring Strategies Symposium, MTS, Washington, DC.

APPENDICES

1. List of Review Panel Members

2. Coastal Monitoring; Background to Gulfwatch Review

3. Summary Table of Recommendations

4. NOAA. 1993. Sampling and Analytical Methods of the National Status and Trends Program, National Benthic Surveillance and Mussel Watch Projects. Vols. 1-4. NOAA Tech. Memo. NOS ORCA 71, Rockville, MD, USA.
(included in Gulfwatch staff copy only)

5. NOAA. 1994. Use of Standards and Reference Materials in the Measurement of Chlorinated Hydrocarbon Residues. NOAA Tech. Memo. NOS OMA 77, Rockville, MD, USA.
(included in Gulfwatch staff copy only)

6. NOAA. 1995. International Mussel Watch Project: initial implementation phase final report. NOAA Tech. Memo. NOS OMA 95, Silver Spring, MD, USA.
(included in Gulfwatch staff copy only)

Appendix 3

EVALUATION OF THE GULFWATCH MONITORING PROGRAM

RECOMMENDATIONS

Recommendations for program modifications made by the review panel for consideration by Gulfwatch staff are not listed here in any priority order. Their implementation will depend on future support levels and the monitoring program priorities that evolve from the present reassessment.

• Revise the draft 5-year retrospective report to consider incorporation of comments made in this review. Focus the report specifically on the Gulf of Maine region and the accomplishments of Gulfwatch.

• Prepare an informative summary report based on the 5-year retrospective for a non-scientific audience; incorporate this activity into the program design.

• Incorporate the preparation of interpretive reports into the program as a regular and essential activity, focusing data interpretation on specific pre-stated hypotheses.

• Assess the results of the initial implementation phase in the context of specific management information needs with the intention of modifying the program to meet these needs. Revisit the original monitoring plan when making this assessment to provide a context for this review, at the time reassessing the current validity of the original plan.

• Assess the results of the initial implementation phase in the context of successful monitoring programs conducted elsewhere with the intention of modifying the program as necessary to meet established "community standards" for coastal monitoring. Incorporate specific review suggestions into this assessment.

• Build on the initial phase experience to create partnerships between Gulfwatch and academic research projects to extend the resources of the program and to establish an on-going informal critique mechanism of program activities.

• Review the existing Gulfwatch QA/QC effort to address QA/QC issues raised by this review. Tightly link an expanded QA effort to the monitoring measurements made and to the reported results; regularly participate in interlaboratory comparison exercises.

• Address the issue of long-term security of Gulfwatch data, data storage and sample archiving.

Appendix 1

EVALUATION OF THE GULFWATCH MONITORING PROGRAM

REVIEW PANEL

BRUCE TRIPP
Asst. Director
Rinehart Coastal Research Center
Woods Hole Oceanographic Institution
MS# 2
Woods Hole, MA 02543-1525
tel. 508-289-2900
fax. 508-457-2172
email <btripp@whoi.edu>

JUDY MCDOWELL
Sr. Scientist
Dept. of Biology
Woods Hole Oceanographic Institution
MS# 2
Woods Hole, MA 02543-1525
tel. 508-289-2557
fax. 508-457-2172
email <jmcdowell@whoi.edu>

JOHN FARRINGTON
Assoc. Director for Education, and

MIKE BOTHNER
Sr. Scientist

Dean of Graduate Studies
Woods Hole Oceanographic Institution
MS# 31
Woods Hole, MA 02543-1541
tel. 508-289-2200
fax. 508-457-2188
email <jfarrington@whoi.edu>

USGS Woods Hole Field Center
Gosnold Lab
Woods Hole, MA 02543-1538
tel. 508-457-2240
fax. 508-457-2309
email <mbothner@usgs.gov>

ANNE GIBLIN
Assoc. Scientist
MBL Ecosystems Center
Water St.
Woods Hole, MA 02543
tel. 508-289-7488
fax. 508-457-1548
email <agiblin@lupine.mbl.edu>

PETER SHELLEY
Sr. Attorney
Conservation Law Foundation
120 Tillson Ave.
Rockland, ME 04841-3416
tel. 207-594-8107
fax. 207-596-7706
email <pshelley@clf.org>

Appendix 2

EVALUATION OF THE GULFWATCH MONITORING PROGRAM

BACKGROUND

Monitoring programs provide a critical link between scientific information and management decisions. Knowledge of the driving biological, chemical and physical processes as well as identification of the sources and quantities of various contaminants is essential to rational resource management decision-making. A continuous, long-term monitoring record is necessary to determine the sources, transport, fate and effects of contaminants and to establish water quality and sediment quality standards and human health standards for the consumption of fish and shellfish.

Monitoring programs for assessing environmental quality should be designed and executed to provide meaningful information on:

(1) the spatial distribution of contaminants;
(2) temporal variability in contaminant distributions; and
(3) the relationship of contaminant inputs to ecological and human health concerns.

Using available but incomplete historic information for the Gulf of Maine, a comprehensive monitoring plan was designed in 1991 that attempted to couple management needs with contemporary understanding of bay-wide processes at time and space scales relevant to the protection of coastal resources. Before agency scientists and managers can begin to monitor the health of coastal waters, the major management issues must first be clearly stated and then testable questions formulated to define the information required to address the management issue

of interest. A technically and fiscally sound monitoring program can then be developed. Such a monitoring program must be designed to generate high quality data in a environmentally relevant context to answer precisely stated management questions. An integrated monitoring program, coordinated among the various resource management agencies, that will enhance our understanding of coastal processes, and improve each agency's ability to make sound resource management decisions is needed. A comprehensive monitoring program that documents changes in the marine system over time will assist in evaluating the success of efforts to abate pollution. Such a program will, of necessity, cut across scientific disciplines as well as political and institutional boundaries.

Resource management problems that face Gulf of Maine agencies are similar to those already identified in other coastal waters (e.g., Capuzzo et al., 1987). Only the mixture and intensity of specific issues change as we move from estuary to estuary. Using available historic information from a variety of previous efforts, major management issues generally include the following:

1. Eutrophication. Excessive enrichment by nutrients from multiple sources, leading to increased plant growth results in declining oxygen levels, changes in benthic community assemblages, and other negative impacts. Eutrophication may have detrimental effects on aesthetics and recreational and fishing activities in coastal waters.
2. Toxics. Poor husbandry of a myriad of synthetic organic, fossil fuel and heavy metal compounds has resulted in widespread contamination of sediment, biota and the water column with implication for living resources and public health.
3. Pathogens. Input of disease causing organisms (bacteria and viruses) to coastal waters is a potential threat to public health and limits access to valuable food resources. A key component of this issue is the need for innovative techniques for assaying pathogens.

The above resource management issues are widely agreed upon (OTA, 1987), but specific testable questions based on these issues must be asked before we begin data gathering. In order to demonstrate that a proposed remedial action is appropriate or that an implemented remedial action is effective, we must be able to detect changes in contaminant concentration and distribution through space and time and distinguish changes resulting form a management action from those that result from natural variation in coastal processes. Examples of management questions are the following:

- What are the spatial and temporal scales and periodicity of anoxia, hypoxia?
- What are the concentrations of specific toxics?
- Are toxics accumulating in commercially import species?
- Is the enumeration of pathogen indicator organisms in ambient waters sufficient to protect public health?

The detection of trends in coastal contamination through space and time is the guiding principle on which a coastal monitoring program is based. In order to detect such trends, sampling must be undertaken with an understanding of natural processes and environmental scales (Farrington et al., 1987). For instance, coastal water masses are driven by storm and tidal energy

and are changing on a minute-to-hour time scale. To provide information of cycling in the water column, water samples must be integrated over a tidal cycle and surface water; spring-fall sampling cannot not represent a water mass. Any monitoring plan must account for the fact that different classes of contaminants (nutrients, toxics, pathogens) each have their own temporal and spatial scales of influence. For example, toxics may affect animal populations through sub lethal effects over long periods of time, whereas long-term reduction of benthic communities due to nutrient loading may be due to short-term anoxic events occurring only every few years. Therefore, any sampling scheme requires scales of measurements appropriate to the scales of variability of the parameter of interest.

In monitoring ecological and human health impacts as a result of contamination of coastal areas, it is important that the environmental objectives of monitoring efforts be defined before sampling is initiated. To understand long-term impacts of chemical contamination in coastal areas, it is important to understand the conditions under which contaminants persist in benthic environments (Farrington and Westall, 1986), the bioavailability of contaminants to commercial resources and the sub lethal effects of contaminants that lead to reduced growth, delayed development, and reduced reproductive effort, with resulting impacts on population stability (Capuzzo and Kester, 1986; Capuzzo et al., 1988). The synergistic effects of complex chemical mixtures must eventually be understood if realistic predictions of contaminant impacts are to be made. Issues such as the potential for deleterious impacts of chemical mixtures on marine species due to increased environmental stress from eutrophication need to be considered as a research topic adjunct to the routine monitoring effort. Ideally, monitoring will be conducted simultaneously with monitoring-related research to ensure that simple monitoring measurements will be interpreted within the context of ecosystem complexities. Our understanding of those ecosystem complexities is at present quite primitive but is continually improving, therefore the links between "monitoring" and "research" must be made and nurtured.

The use of sentinel organisms (i.e., mussel watch) is one valid approach to coastal monitoring (Farrington et al, 1983; NOAA, 1984, 87, 89, 91, 91a). It is not the only approach and it is not the best approach for some management questions. The use of bivalves is a tried and true method for coastal monitoring but it is only one tool in a toolbox filled with alternative monitoring techniques. By analogy, a medical doctor will measure body temperature as an indicator of general health, but will use many other tests as he/she attempts to diagnose a specific disease. Any good monitoring program will keep this point in mind as conditions change and as reassessments are made.