Prototype System for Multidisciplinary Shared Cyberinfrastructure: Chesapeake Bay Environmental Observatory

CBEO Project Team

Abstract: A prototype system for a Chesapeake Bay Environmental Observatory (CBEO) is under development by a multidisciplinary team of researchers from the domains of environmental engineering, marine science, hydrology, ecology, and computer science. The vision is to provide new means of coupling and synthesizing field sampled and model generated data in a way that will open up new data sources to researchers and managers interested in understanding and resolving some currently unanswered questions and problems concerning hypoxia in the Chesapeake Bay. It will do so by developing advanced cyberinfrastructure to provide uniform nationwide access to new tools and a wide variety of data of disparate type, scale, and resolution, in both spatial and temporal domains, including model-derived data from past runs of major computational models for Chesapeake Bay hydrodynamics and water quality. Some key goals of the prototype project are to resolve the existing data source heterogeneities such that all relevant data are accessible through one interface, to archive and facilitate the analysis of model input and output files, and to provide new shared tools for data analysis, all with the goal of transforming the way scientific research and science-based management is conducted on the Chesapeake Bay. There are four project teams operating separately but in close and continuous communication. The teams' objectives are to make simultaneous and parallel advances in (1) environmental observatory network design and nationwide network access to Bay data (CBEO:N); (2) furthering the educational missions and outreach at the host institutions (CBEO:E); (3) providing a test-bed application that will allow the development and testing of new cyberinfrastructure and data analysis tools (CBEO:T); and (4) using all of the above-mentioned advances on focused science questions to demonstrate the transformative nature of the CBEO for addressing research questions and improving

[1]William P. Ball, Professor and CBEO Project Co-Director, Dept. of Geography and Environmental Engineering, Johns Hopkins Univ., 3400 North Charles Street, Baltimore, MD 21218. E-mail: bball@jhu.edu

[2]Damian C. Brady, Postdoctoral Associate and CBEO Test-Bed Team Member, Dept. of Civil & Environmental Engineering, Univ. of Delaware, 301 DuPont Hall, Newark, DE 19716. E-mail: damian@udel.edu

[3]Maureen T. Brooks, Faculty Research Assistant and CBEO Science Team Member, Univ. of Maryland Center for Environmental Science - Horn Point Laboratory, 2020 Horns Point Rd., Cambridge, MD 21613. E-mail: mbrooks@hpl.umces.edu

[4]Randal Burns, Assistant Professor and CBEO Test-Bed Team Leader, Dept. of Computer Science, Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21218. E-mail: randal@cs.jhu.edu

[5]Benjamin E. Cuker, Professor and CBEO Education and Outreach Team Leader, Dept. of Marine and Environnemental Science, Hampton Univ., Hampton, VA 23668. E-mail: benjamin.cuker@hamptonu.edu

[6]Dominic M. Di Toro, Professor and CBEO Project Director, Dept. of Civil and Environmental Engineering, Univ. of Delaware, 345 DuPont Hall, Newark, DE 19716. E-mail: dditoro@udel.edu

[7]Thomas F. Gross, Program Specialist, UNESCO, 1, rue Miollis, 75732 Paris, France; formerly Administrative Director for CBEO Project and Research Scientist with Chesapeake Research Consortium, Chesapeake Community Modeling Program, Edgewater, MD 21037. E-mail: t.gross@unesco.org

[8]W. Michael Kemp, Professor and CBEO Science Team Leader, Univ. of Maryland Center for Environmental Science-Horn Point Laboratory, 2020 Horns Point Rd., Cambridge, MD 21613. E-mail: kemp@hpl.umces.edu

[9]Laura Murray, Professor and CBEO Education and Outreach Team Leader, Univ. of Maryland Center for Environmental Science—Horn Point Laboratory, 2020 Horns Point Rd., Cambridge, MD 21613. E-mail: murray@hpl.umces.edu

[10]Rebecca R. Murphy, Graduate Student and CBEO Test-Bed Team Member, Dept. of Geography and Environmental Engineering, Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21218. E-mail: rebecca.murphy@jhu.edu

[11]Eric Perlman, Graduate Student and CBEO Test-Bed Team Member, Dept. of Computer Science, Johns Hopkins Univ., 3400 North Charles St., Baltimore, MD 21218. E-mail: eric@cs.jhu.edu

[12]Michael Piasecki, Associate Professor and CBEO Network Team Leader, Dept. of Civil, Architectural, and Environmental Engineering, Drexel Univ., 3141 Chestnut St., Philadelphia, PA 19104. E-mail: michael.piasecki@drexel.edu

[13]Jeremy M. Testa, Graduate Student and CBEO Science Team Member, Univ. of Maryland Center for Environmental Science-Horn Point Laboratory, 2020 Horns Point Rd., Cambridge, MD 21613. E-mail: jtesta@hpl.umces.edu

[14]Ilya Zaslavsky, SDSC Spatial Information Systems Laboratory Director and CBEO Network Team Leader, San Diego Supercomputer Center, Univ. of California San Diego, 9500 Gilman Dr., La Jolla, CA 92093–0505. E-mail: zaslavsk@sdsc.edu

Note. Discussion open until March 1, 2009. Separate discussions must be submitted for individual papers. The manuscript for this paper was submitted for review and possible publication on June 14, 2007; approved on December 5, 2007. This paper is part of the ***Journal of Hydrologic Engineering***, Vol. 13, No. 10, October 1, 2008. ©ASCE, ISSN 1084-0699/2008/10-960–970/$25.00.

management approaches for large coastal systems that are heavily affected by humans. Finally, this CBEO cyberinfrastructure development is geared toward ensuring that the envisioned system is integrated into larger nationwide environmental observatory network initiatives (e.g., Water and Environmental Research System, Ocean Research Interactive Observatory Networks, National Ecological Observatory Network, and Long Term Ecological Research) thus helping to lead the way toward the development of a continental-scale environmental observatory network.

DOI:

CE Database subject headings: Chesapeake Bay; Environmental engineering; Hydrodynamics; Research; Internet; Drainage; Ecology.

Motivation and Overview

The use of large-scale environmental observatory systems to conduct transformative research has received much attention in the environmental science and engineering community recently. The National Science Foundation (NSF) is currently supporting the development of environmental observatory prototypes with the aim of building a substantive knowledge base for conceiving and building these systems. Some of NSF's environmental observatory initiatives include the Consortium for the Advancement of Hydrologic Science (CUAHSI 2007), the Collaborative Large-scale Engineering Analysis Network for Environmental Research (CLEANER 2007), both of which are now combined to form the Water Environmental Research Systems (WATERS) network, the National Ecological Observatory Network (NEON 2007), and the Ocean Research Interactive Observatory Networks (OOI 2008). Within this context, the NSF solicited proposals in late 2005 for projects to develop and deploy a "prototype cyberinfrastructure for environmental observatories" [CEO:P] and then "demonstrate [its] viability in order to inform the planning of, and development of, an environmental cyberinfrastructure for large-scale, environmental observing systems" (NSF 2005).

The Chesapeake Bay Environmental Observatory (CBEO) is currently being designed by the multidisciplinary CBEO project team as a prototype to demonstrate the potential of cyberinfrastructure for transforming environmental research, education, and management. Because the Chesapeake Bay is a large environmental system that is host to numerous ecological, environmental engineering, hydrologic, coastal, marine, and social complexities, it is ideally suited to serve as a prototype test bed designed for a large and very diverse user community. In fact, it is envisioned that the system will later be integrated to multiple environmental observatory networks, as schematically illustrated in Fig. 1. As a result, this project pursues an end-to-end approach to two major components of cyberinfrastructure, i.e., a test-bed observatory and a new type of node for an existing national network. These components will be useful for both regional applications and for shared cyberinfrastructure with other researchers nationwide.

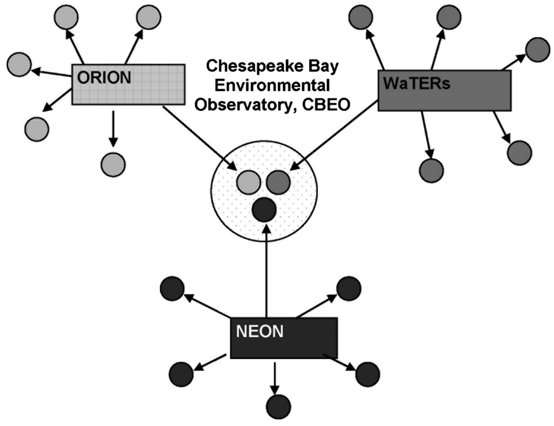

Fig. 1. Conceptual integration of CBEO into the other three large scale environmental observatories

The CBEO is under development by a multidisciplinary project team that is led by PIs comprising marine/ocean scientists (Cuker, Gross, Kemp, Murray), computer scientists (Burns, Zaslavsky), and environmental engineers (Ball, DiToro, Piasecki). The NSF required that the observatory be designed and built using existing cyberinfrastructure components from other NSF programs. Toward this end, the San Diego Supercomputer Center, which has deployed software components of the NSF-sponsored GeoSciences Network (Baru 2004), is working as part of the CBEO team to implement and develop a CBEO portal on this network. The CBEO project not only uses software stacks and other approaches developed for the GEON web interface, but also integrates metadata standards developed through CUAHSI-Hydrologic Information Systems/WATERS and applies data federation and querying approaches developed for the National Virtual Observatory (Szalay et al. 2002; Malik et al. 2003). Thus, the prototype components developed in this work are assured to be flexible and amenable to extension and upgrade.

Finally, a key objective was that the project "works with real environmental data, includes a mechanism for general users to access the prototype, and demonstrates the utility of its approach by attracting users outside of those researchers directly involved in the project." In this context, the CBEO team has targeted the seasonal depletion of dissolved oxygen, or hypoxia, in the Chesapeake Bay as a fundamental issue around which to develop specific environmental engineering and science questions that will help to structure, test, and demonstrate the prototype's capabilities. The use of seasonal hypoxia as a key environmental problem is well suited because of the long-standing history of hypoxia research in the Bay and the general importance of this issue nationwide. In addition, the likelihood is high that it will provide answers to key questions about coastal hypoxia that will lie at the intersections of existing observational and modeling data sets. A key premise of the work is that cyberinfrastructure and information technology are now at a point where new tools can be effectively developed and deployed to better link, analyze, and visualize multiple observational and modeling data sets of disparate size, scale, and resolution (in both spatial and temporal aspects) that will transform our ability to address large, policy-oriented science questions. The ability of cyberinfrastructure to perform such functions, however, has yet to be properly explored and demonstrated. It is our hope that the CBEO and other prototypical projects and test beds will be able to provide the needed research and development for such a demonstration.

The CBEO prototype development effort was motivated by a

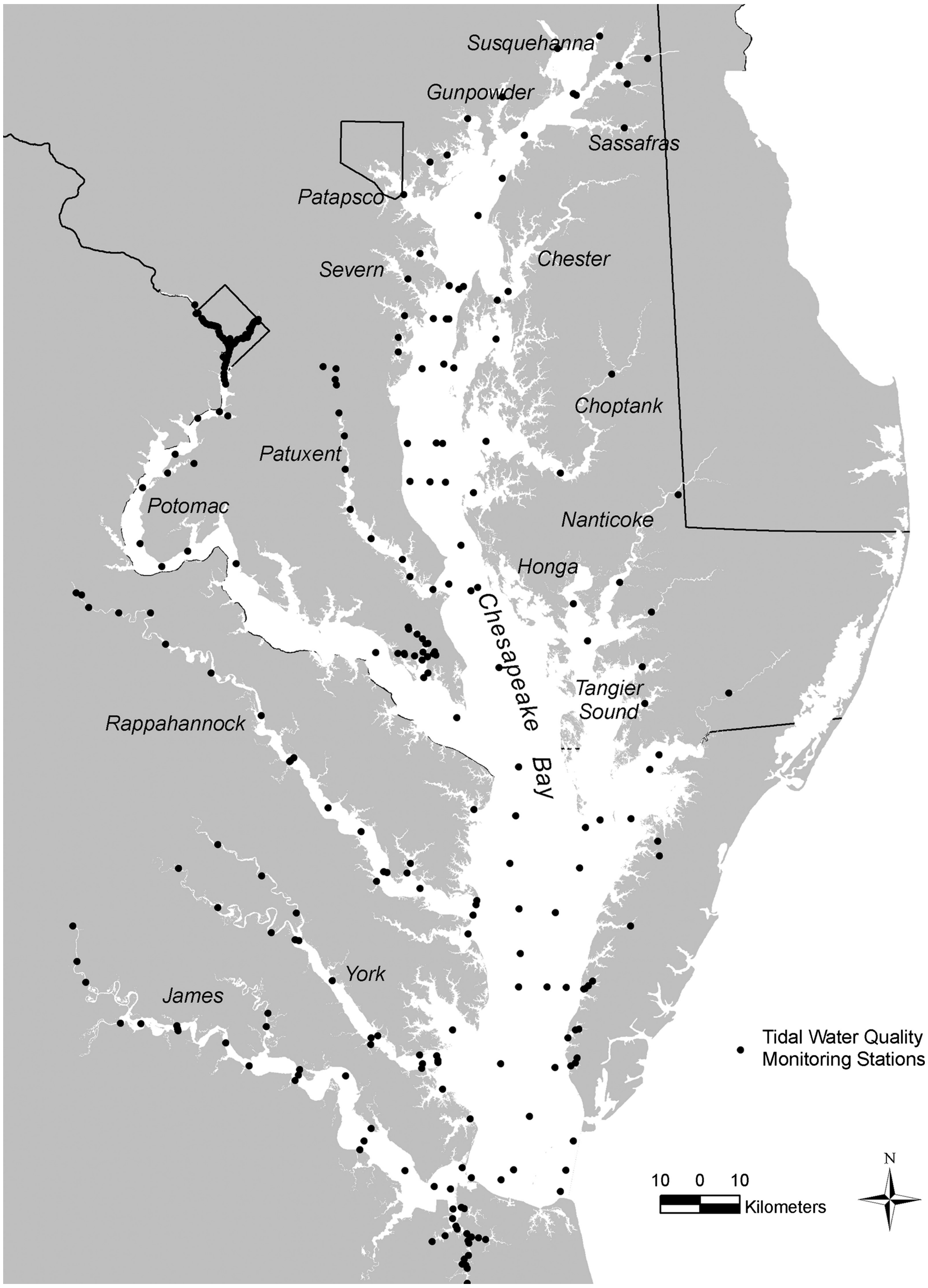

Fig. 2. Fixed stations for tidal water quality monitoring under the Chesapeake Bay Tidal Monitoring Program. Monitoring has been conducted at multiple depths once or twice monthly (less frequently in winter) since 1985 and at some stations for much longer. (Reprinted with permission from the U.S. EPA Chesapeake Bay Program Office, Annapolis, Md., 2007.)

hypothesis that the development of new and better tools for finding, viewing, and analyzing multiple data sets and data streams, in conjunction and comparison with each other and with model-derived data, should lead to new knowledge about Chesapeake Bay hypoxia and toward better understanding of management solutions. This manuscript provides a summary of the project being undertaken to test this hypothesis, with a focus on the design challenges and project approach.

Science and Cyberinfrastructure Challenges on the Chesapeake Bay

Chesapeake Bay as a Study Site

The Chesapeake Bay (CB, Fig. 2) is an ideal location for researchers to demonstrate how cyberinfrastructure can be used to help answer complex and unresolved science questions, enhance the ability of educators to teach environmental science, and inform management of environmental resources. The Chesapeake is a well-studied and intensely monitored coastal ecosystem with active research and resource management activities. The degradation of water quality, seagrass communities, and benthic animal populations from human activities is well documented (e.g., Kemp et al. 2005). The region has a long-standing history of collaboration between researchers and managers, with the objective of basing management on sound scientific understanding of system processes and responses to disturbance (e.g., Malone et al. 1993; Boesch et al. 2001). Federal mandates and multijurisdictional initiatives have motivated an aggressive Bay restoration program (Boesch et al. 2001). A systematic monitoring program has been in place since 1984, and numerous integrated research programs have assessed various aspects of system dynamics (e.g., Kemp et al. 2004, 2005; Roman et al. 2005). In addition, the Chesapeake Bay management program has supported development of a coupled hydrodynamic and water quality model that has been calibrated (Cerco and Cole 1993; Cerco 1995; Johnson et al. 1993), upgraded (e.g., Cerco and Noel 2004), and repeatedly run to produce multiyear (1986–2000) simulations of physical circulation and ecological dynamics, at spatial scales of 10 to 1 km^2 horizontal and 1 to 5 m vertical and temporal scales of minutes to hours (Cerco 1995). These and other models are available to all researchers and managers, access to which is facilitated by the Chesapeake Community Modeling Program managed through the Chesapeake Research Consortium (CRC et al. 2008).

In addition, there is a wealth of collected and sampled data, a wide variety of incoming data streams, and a very extensive numerically generated data. Some of the existing data and data streams have already been collected and organized by the U.S. EPA Chesapeake Bay Program Office (http://www.epa.gov/region03/chesapeake/; http://www.chesapeakebay.net/), where they have been made available for public access through the Chesapeake Information Management System (CIMS). This system provides a rich source of information that is already available for enhanced use through the development of new cyberinfrastructure tools. The vision and challenge for the CBEO is to demonstrate that cyberinfrastructure can provide hypoxia researchers, managers, and educators with transformative new tools for synthesizing, harvesting, analyzing, and interpreting spatially and temporally distributed information of unprecedented diversity and density.

Chesapeake Bay Science: Challenges, Questions, and Resources

Hypoxia as a Prototypical Challenge for Scientific Research

The seasonal depletion of dissolved oxygen (O_2) from coastal waters, i.e., hypoxia, is a widespread problem of growing proportions that is tied to human influences (Rabalais et al. 2002; Howarth et al. 2000) and is of worldwide concern (e.g., Rosenberg 1990; Diaz 2001). Accumulating evidence indicates that the intensity and extent of summer hypoxia in Chesapeake Bay have been increasing since the early 1950s (Hagy et al. 2004). Although *interannual* variations in extent and duration correlate with fluctuations in freshwater discharge [Fig. 3(a) upper panel], *long-term* increases follow broad trends of increasing nutrient loading from the watershed [Fig. 3(a) lower panel, Hagy et al. 2004]. Thus, hypoxia in coastal waters is a consequence of interactions between watershed land-use (and associated nutrient loading), hydrology, climate, and oceanographic processes. An

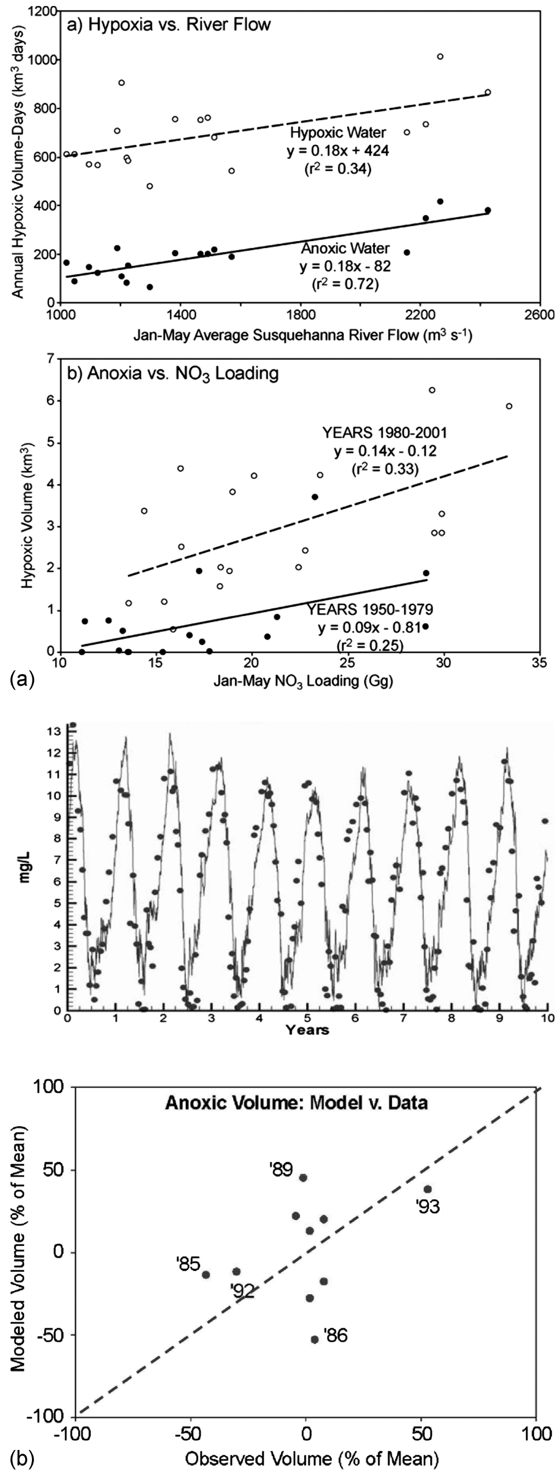

Fig. 3. (a) Relationships between bay hypoxia and river flow or nutrient loading. (a) Time-integrated volume of hypoxic and anoxic (<0.2 mg O_2 l^{-1}) water versus winter-spring river flow. (b) Mid-summer volume of anoxic water versus nitrate loading for earlier (1950–1979) and more recent years (1980–2001) (figure adapted from Hagy et al. 2004). (b) Simulations of a water quality model for Chesapeake Bay (Cerco 1995) showing (a) close match for bottom dissolved oxygen concentrations at seasonal scales for one location (CB5.2), but (b) relatively poor match for interannual variations in the integrated volume of summer hypoxic water (C. Cerco, communication).

environmental observatory could transform our process understanding, because some fundamental and unanswered questions about hypoxia remain (Nixon 1995, Cloern 2001). Our ability to understand and address these scientific questions would be enhanced in transformative ways with vastly improved methods for finding, integrating, interpolating, and visualizing multiple data sets (including model output) that are complementary but disparate in their structure, precision and scale (resolution and extent).

Chesapeake Bay Science Questions

The team identified three related questions pertaining to hypoxia. The first derives from Fig. 3(a) (lower panel), which indicates that for a given rate of N loading (total N is highly correlated to NO_3), a larger hypoxic volume has generally been observed during the last 20 years in contrast to the previous three decades (1950–1979). The considerable scatter in the relationships is largely attributable to fluctuations in river flow. If proven significant, these differences would imply the existence of highly nonlinear underlying mechanisms that are currently not well understood but which have major implications for prediction and management (Kemp et al. 2005). Conclusively validating these trends and elucidating their causes would transform our understanding of the processes causing hypoxia.

The second question pertains to the observed relationships between year-to-year variations in hypoxia and river flow [e.g., Fig. 3(a), upper panel]. Close examination of these trends reveals considerable scatter which is poorly understood, and it is unclear why the existing coupled hydrodynamic-water quality model, although calibrated to reproduce seasonal patterns [Fig. 3(b)], is incapable of simulating these interannual trends [Fig. 3(b)]. We suspect that there are problems both with estimating the time course of hypoxic volume and with modeled processes, both hydrodynamic and biological. Exploring these issues requires better integration of existing observational data sets with each other and with the model-derived output.

The third question involves the regional and seasonal balance between source and sinks of organic matter that regulate O_2 consumption and hypoxia in bottom waters. Although prior analyses suggested that organic matter production by algae in shallow waters may drive respiration in deep channel waters (Kemp et al. 1997), the general structure of the water quality model being used for management (Cerco and Noel 2004) reflects the more traditional view that phytoplankton production in surface waters of the deeper regions of the Bay is the major source of organic input fueling hypoxia (e.g., Malone et al. 1988). A recently established shallow water monitoring program should provide the requisite data (MD DNR 2005) that will be incorporated into the CBEO.

Data Resources for the Chesapeake Bay

Varied observatories and monitoring programs generate numerous Chesapeake Bay databases relevant to hypoxia including water quality conditions, physical structure and circulation, sediment characteristics, biogeochemical processes, abundances of plankton, benthos, and fish populations. Extensive sets of data have been and are being collected at a wide range of scales using diverse means including grab samples, moored and towed sensor systems, and satellite and aircraft remote sensing. Additionally, data are available from other disparate sources: acoustic sampling of water depth, sediment character and fish abundance; quantitative and qualitative chemical and physical data from sediment cores that have been used to recreate ecological history and to measure the rates of biogeochemical processes (e.g., Cooper and Brush 1991); and output data for and from a wide variety of computer models (Table 1). Overall, the time and space scales of the available data range from minutes to decades and from centimeters to hundreds of kilometers.

An important goal of environmental observatories is to make such disparate data sources accessible in a user-friendly way and on a web-based workbench that also provides the necessary tools for properly resolving and interpolating their different spatial and temporal scales and visualizing the results. The realization of this goal would transform our ability to understand and interpret the data. For example, Fig. 4 shows fine-scale distributions of water quality obtained from towed undulating sensor systems throughout the Bay. Data from this monitoring program include measurements of physical, chemical, and biotic aspects of water quality, but the complementary large databases with key ecological rates (e.g., primary productivity, community respiration and nutrient recycling, and organic particle sinking) were obtained under different research programs at different scales and times. Moreover, output of numerical model computations provides physical circulation, water quality, and ecological properties at a fine resolution. To address the scientific questions posed, we must interpolate and integrate the various databases so that computations of flux, transformation, and coherence can be made and patterns as well as mechanistic relationships can be visualized and quantitatively analyzed.

The CBEO team is focusing on the use of specific hypoxia-relevant data sets from Table 1 to provide new insights into specific science questions. Through parallel science activities on a local server containing these multiple data sets (our test bed) and through cyberinfrastructure activities that are designed to make these data amenable to analysis over the network, we intend to demonstrate the scientific value of combining disparate data with new tools, and simultaneously develop the cyberinfrastructure needed for better sharing such data and tools to other users. Overall, this project aims to use prototypical science questions, data, and innovative new approaches to obtain insight into the intrinsic value of cyberinfrastructure for research.

Challenges and Objectives for Shared Cyberinfrastructure on the Chesapeake

Resolving problems of structural, syntactic, and semantic heterogeneity across data sets maintained at different locations for different purposes is a major challenge facing all environmental observatories. In addition, there are challenges relating to problems of semantic and syntactic heterogeneities among data sets, lack of relevant metadata descriptions for some data, and the major technical challenge of integrating data collected at different spatial and temporal resolutions, under different sampling schemes, with different frequencies and extents of coverage, and with different levels of precision and uncertainty. Formal representations and metadata for such disparate data sets are required. This is a prerequisite for developing scalable and robust means for querying, exploring, and integrating observation data, and preparing them for further analyses. The CBEO project will address these challenges by developing data registration, query, integration, and analysis tools for shared use within the CBEO test bed as well as across multiple networks through the portal (Fig. 1)

Inconsistent Semantic and Syntactic Representations

A major challenge in the federation of disparate data sets is developing and implementing appropriate metadata standards and data registration systems that fully reflect data properties. The key to data interoperability at a basic level (search, viewing, retrieval,

Table 1. Space and Time Scales of Chesapeake Bay Databases Relevant to Hypoxia Questions

Database	Variables included	Space scales[a] Grain	Space scales[a] Extent	Time scales Grain	Time scales Extent (years)	Web site address or data source
Grab sample monitoring	T, S, Chl, N, P, C, Si, O_2, k_d	z, 1 m x, 10 km	0–30 m 300 km	2–4 weaks	20	www.chesapeakebay.net
Fixed sensors	T, S, Chl, O_2, k_d	0 km	1–10 km	15 min	2–5	www.hpl.umces.cbos.edu
• Shallow (DNR) • Deep (CBOS)	T, S, u, v, w	10 km	300 km	15 min	5	www.chesapeakebay.net www.cbos.org
Underway sensors	T, S, Chl, O_2, k_d	1 m	1–10 km	4 weeks	1–3	www.eyesonthebay.net
• Shallow (Dataflow) • Deep (NSF TIES)	T, S, Chl, O_2, zooplankton	z, 1 m x, 10 km	10 m 100 km	3 months	5–6	www.chesapeake.org/ties Roman et al. 2005
Remote sensing	Chl	1–10 km	200 km	1–10 day	10	Harding et al. 1994
• SeaWiFS (satellite) • ODAS (aircraft)	Chl	5 m × 50 m	300 km	1–4 weeks	10	Harding et al. 2005
Bathymetry	Depth, volume	z, 1 m, x, 1 km	300 km	na	na	Cronin and Pritchard 1975
River inputs	Flow	10 km	300 km	1 day	50	www.usgs.gov
Nutrient load	N, P, Si	10 km	300 km	1 day	20	www.chesapeakebay.net
Climate	T, rainfall, wind, tides	10 km	300 km	1–24 h	50	www.noaa.gov
WQ model	T, S, Chl, N, P C, Si, O_2, k_d	z, 1 m, x, 1 km	300 km	1 h	15	C. Cerco (US Army CoE), private communication
Hydro. model	T, S,u, v, w, $K_{x,y,z}$	z, 1 m, x, 1 km	300 km	5 min	15	B. Johnson (US CoE), private communication

Note: T=temperature; S=salinity; Chl=phytoplankton chlorophyll-*a*; O_2=dissolved oxygen; N=nitrogen; P=phosphorus; C=organic carbon; Si=silica; k_d=diffuse light attenuation coefficient; u, v, w=velocity components in x-, y-, z-directions; $K_{x,y,z}$=turbulent mixing in x, y, z.

[a]Cartesian coordinates: "x" follows land–sea gradient; "y"=horizontal axis perpendicular to x; and "z"=vertical axis.

and analysis) is resolution of syntactic and semantic differences through a mediation layer (Ludäscher et al. 2007). This layer reconciles the different metadata standards and implements vocabularies that allow connections between otherwise disparate data descriptions, e.g. "Gauge Height"="Stage." Although this example is trivial, similar inconsistencies exist in the way metadata annotations are encoded (plain text versus XML) and published, and the manner in which data are stored (location, format). This is a serious obstacle to the interoperability of environmental observatory networks, because the ecological community networks are either already using Long Term Ecological Research, or planning to adopt (NEON) the Ecological Metadata Language (EML 2008) for metadata descriptions, whereas the ocean community network (OOI 2008) is slated to use Integrated Ocean Observing System (IOOS)-developed [Federal Geographic Data Committee (FGDC)-based] Data Management and Communications (OCEAN.us 2005) and the WATERS community [through the Consortium of Universities for the Advancement of Hydrologic Sciences, Inc., hydrologic information system (CUAHSI-HIS)] is planning to use the International Standards Organization (ISO)-based standard 19115 (ISO 2007). First experiences with the Chesapeake Bay Information Management System (CIMS) and its metadata annotations have revealed two important findings. First, in cases where data providers have adopted a metadata standard to annotate their data, this establishes a relatively straightforward means to develop a formalized metadata framework that can be used for crosswalks. The writers have now had experience in mapping the provided metadata (CIMS is mostly FGDC based) to a structure based on the ISO 19115 and have found little difficulty in mapping the metadata tags—see also the FGDC-based North American Profile for the ISO 19115 (FGDC 2007). The second finding, which is more important in the writers' view, is that there has been some difficulty in finding consistent vocabularies that can be properly mapped and also difficulty in reconciling the fact that some metadata tags in one system have no equivalent in the other. This often means that some metadata may be lost when mapping from an expressive system to one that requires less entries, or having to create metadata annotations when trying to adopt from a sparsely annotated system. First experiences have shown that the map-

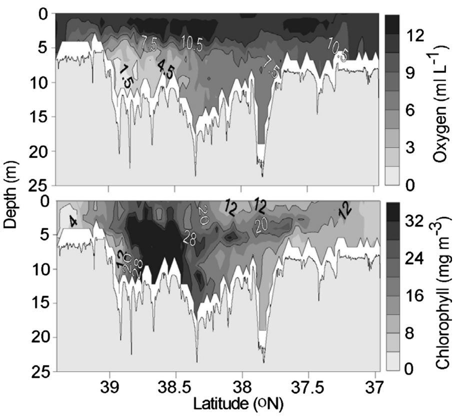

Fig. 4. Dissolved oxygen (upper panel) and chlorophyll-*a* (lower) derived from transects with a vertically undulating sensor system (*Scanfish*) towed along the central axis of Chesapeake Bay (April 2000). Data from the NSF LMER:TIES project (Table 1; see Roman et al. 2005).

pings work out relatively easily for most of the CIMS data, although the comparatively greater richness of the CIMS metadata structure did provide some challenges. In this regard, the CUAHSI-HIS-based system only requires about 36 annotations per data value (or set), which are very basic in their structure (CUAHSI 2007). Higher degrees of interoperability will therefore require that metadata tags be mapped, semantics reconciled, and both syntactics and semantics implemented in a way that allows automatic mediation without human intervention (e.g., Bermudez and Piasecki 2006).

Integration of Disparate Information Systems

Integration of data collected within heterogeneous but overlapping cross-disciplinary observation networks requires that the CBEO rest on a sound but flexible technological foundation, e.g., that established by the GeoSciences Network, GEON (Baru 2004). Required components for data integration include: (1) format, projection, and unit conversion; (2) knowledge-based data registration and query rewriting, and (3) spatiotemporal interpolation techniques. The most important challenges are to develop a framework that (1) can be modified and extended; (2) is sufficiently flexible, transparent, and documented to allow integration into future new systems; (3) is based on well-established standards with a high degree of provenance; (4) has an access interface that can be programmed against; and (5) is robust and scalable for the supported data sources, data types, and interpolation models. Observatory developers must find conceptual representations that describe both the storage structure (i.e., database and file system) and the descriptive elements (e.g., metadata tags, semantic conventions). It is also important to permit basic access to data through web services that are operating system independent, can be integrated into multiple user applications and workflows, and can use a uniform simple logic. Web services that provide data exposure to the "outside" world are one of the great promises of the Water Environmental Research System (WATERS) and other observatory networks and substantial progress is being made. For example, SOAP web services that permit access to USGS National Water Information System have been made public through nationally supported servers for use in custom desktop or web-based end-user applications, such as the Data Access System for Hydrology (CUAHSI 2007; Maidment et al., 2006). The CBEO is building on technologies developed under both the GEON and CUAHSI-HIS programs. In particular, the newly developed CBEO portal (http://geon16.sdsc.edu:8080/gridsphere) is based on GEON technologies, and supports registration, discovery, and integration of data resources of different types, contributed by CBEO members as well as by the wider earth sciences community. Access to observation data for the Chesapeake Bay will be provided via a Workgroup HIS server, a node in the emerging system of hydrologic observation servers being developed within the CUAHSI-HIS project.

Merging of Data for User Applications

New means will be needed to integrate data that are collected at disparate spatial and temporal scales, and new approaches will be needed to merge observational data and model-derived results. Many Bay-related data sets fall into two classes: four-dimensional data (4D=3 spatial+1 time). If variable type (e.g., temperature, velocity, concentrations) is also considered as a dimension, then the data are five dimensional (5D). In this context, a metadata description is also an additional variable that is useful for such tasks as "weighting" other types of data for display, interpolation, or averaging. Thus, either the value of nitrate concentration or the metadata about nitrate data quality can be viewed as the "variable" that comprises the fifth dimension.

Moreover, for any given variable, many types of data do not occur at single locations in space and time, but are rather association with polygons in 4D space–time and represent averages of some sort, e.g., a concentration derived from a photograph pixel or a measured (model-derived) concentration within a sampled (grid element) volume over the compositing time of the sample (or temporal discretization of the model). Another example would be water velocity across some geometrically defined interface. The rest of the data universe can be even more complex, such as descriptive data for sediment cores and living resources.

The CBEO project is currently focusing on 5D observational water quality data and the development of better techniques for comparing among and between different kinds of observational and model-generated data. We are already facing challenges of how best to store, retrieve, and make effective use of the additional polygonal information relevant to model-generated results and pixilated data sets. Storage and use of other (even more descriptive) types of data (such as that from biological sampling and sediment cores) will represent even greater challenges, but are beyond the scope of our current prototype.

Configurability of a Virtual Observatory

The CBEO will be a multipurpose and multifaceted observatory. However, each user's individual view should meet their particular needs. Such a custom CBEO view will likely represent a particular spatiotemporal window over a subset of observatory data sets and available processing workflows. Several existing cyberinfrastructure projects allow for such personal user areas maintained at grid nodes, such as (myGEON 2006) in the GEON project, and myLEAD in the LEADPortal (2008) project. In the personal areas, authorized users may reference data sets and other resources found in resource catalogs and perform supported functions using common software stacks that include semantic annotation, data transformation, and online mapping. In GEON, an ability to develop user-defined processing workflows is also being added.

In the proposed system, we will extend the personal workbench concept to enable users to define an integrated view over selected CBEO resources, specify how these resources are linked, and instantiate the symbolic representation as a standards-compliant XML document or a well-known format with accompanying metadata, which in turn can be used as input to analysis programs. The latter step would rely on GEON or Scientific Environment for Environmental Knowledge services for resolving semantic mismatches.

Challenges for Interpolation

Integrating and analyzing relationships between data sets is difficult because of their spatial and temporal heterogeneity. Unlike most database applications, data cannot be joined by equality or a simple predicate. Rather, data have complex relationships defined by the underlying science and properties of data generation, e.g., sampling discipline and measurement error. Generally, putting the data on a common basis requires interpolation algorithms. For example, a scientific question that examines point data collected hourly at buoys with pixilated data collected daily by satellite requires interpolation in space–time.

There are two important and challenging aspects to data interpolation. First, the distributed nature of the CBEO complicates the evaluation of interpolation functions. Data can be brought together at a single site and interpolated locally. In a distributed environment, this strategy requires transfer of unnecessarily large amounts of data, e.g., all readings used in the interpolation kernel. For performance reasons, interpolation should be evaluated at the databases resulting in smaller data transfers (Malik et al. 2005) thus improving performance. To allow scientists to customize retrieval and interpolation functions, we propose to develop interpolated spatial joins using SQL language extensions and user-defined functions. User-defined functions allow complex programs written in Java, C++ and C# to be executed at the database (Chamberlain 1998). They are the fundamental construct that enables database federations to "bring the computation to the data" (Szalay et al. 2002). The challenge here is to provide class libraries of interpolation functions that can be invoked directly and provide customized interpolation functions to be executed at the database.

Second, a common feature of all observational data is its incompleteness in 4D, in contrast to model derived data that are continuous in time and exist in all polyhedral model segments. The problem of commonly used interpolation schemes, from simple weighting schemes, e.g., inverse least squares, to more complex methods, e.g., kriging, is the need for an interpolation metric to be used in relating observed to interpolated points. In regard to kriging, for example, ordinary kriging based on location may not be sufficient, as the system is heavily influenced by land locations, density gradients, and other conditions that influence the movement of water and connectivity among data points.

The presence of model data, specifically hydrodynamic flow fields, as part of CBEO suggests that an interpolation scheme can be devised that uses this information. This method has been used with flow fields that are estimated from the data to be interpolated (Ueng and Wang 2005; Yang and Parvin, 2003; Zhang and Akambhamettu 2003). Model-derived flow fields from many years of hydrodynamic model simulations of the Chesapeake Bay (e.g., Cerco 1995) are available to the CBEO. These are independent of the data and can therefore be used with sparse observational data. In theory, barriers to transport such as pycnoclines and land boundaries could be properly taken into account through the effective use of accurate hydrodynamic information, thus expanding our ability to view and interpret sparse observations. One important challenge for the CBEO project team is to demonstrate that our cyber-based observatory can in fact facilitate the development and application of tools of this type.

Challenges from a User's Perspective

From a user's point of view, the CBEO should be the vehicle for performing research using disparate collections of data sets and model outputs that could not previously have been used together. In the context of hypoxia, a user might be most interested in applying merged 5D data of the type discussed earlier. The following are examples in the form of query scenarios.

1. **Queries for Existence and Extent of Data.** For a variable in a particular spatial and temporal domain, with a defined granularity, what is the data coverage? CBEO would return a 5D data cube representing presence/absence of data that for visualization purposes could be subsampled: one variable at a time (4D), grouped into time intervals (three dimensional), and spatially sliced and collapsed into two-dimensional slices.
2. **Queries for a Data Merge.** Select from an extant 5D subset and produce the merged data set. Select an averaging element in 4D space–time. Populate all the elements with point estimates or multiple entries if overlapping data are present. Preserve the data polygonal structures. Visualize the results using tools described in 1.
3. **Queries for Data Interpolation.** For filled elements, specify compositing algorithms in time and space (e.g., volumetrically weighted averages that incorporate point values). For unfilled elements, choose an interpolation algorithm, for example, through universal kriging with a variety of alternative covariates.
4. **Queries for Analysis.** Possibilities are numerous. For example: (1) *Space–time averages*: the volume of the space–time isosurface for $O_2 < 2.0$ mg/L; (2) *Areal mass fluxes*: daily average flux of nitrate across a plane in the bay. Flux of O_2 across the pycnocline, defined as a surface bisecting, e.g., a 1.0 g/L m vertical change in salinity; (3) *Sources*: the net monthly production of algal biomass in the photosynthetic zone accounting for loss by turbulent mixing and settling to the sediment or: the sediment flux of ammonia during oxic and anoxic periods.
5. **Queries for Model–Data Comparison.** A typical annual model run can produce 100 Gbytes of output for the hydrodynamics and the 25 computed biological and chemical variables. During the same year there are perhaps 1 Gbyte of observational data. The CBEO can be queried in order to compute and visualize statistics at various time and space scales.
6. **Queries for Data Assimilation.** Produce a "source difference" 5D cubes in which the fifth dimension is the source/sink that must have been present to produce the observed concentration, corrected for transport effects using the velocity and mixing data. Arguably, this is the most useful form of data assimilation as the output is a direct quantification of the processes responsible for the changes in the state variables.

These six query scenarios are the basis for web services and user interface development targeting the following three categories of users: (1) project PIs and their research teams, for whom configurability and flexibility of the CBEO environment and the ability to quickly integrate data sets are paramount; (2) undergraduate and graduate students, with the emphasis on observatory data exploration, packaged queries, and curriculum integration; and (3) researchers and system integrators outside the observatory domain, for whom CBEO shall provide a gateway to well-documented Chesapeake Bay observations and model-derived data.

Project Approach: CBEO Development

To better manage and assign the various tasks the project team has adopted a framework called NETS. NETS divides the CBEO scope into for four subtopics, i.e., (N)etwork, (E)ducation, (T)est bed, and (S)cience. The CBEO:T and CBEO:N teams are working with selected databases, data streams, and model-derived data from the extensive body of existing data to provide an access mechanism for outside users. The CBEO:E team is responsible for implementation and dissemination of material that is conducive for K–12 education as well as higher education. The CBEO:S team formulates the science questions and provides feedback to the test-bed and network teams for improving and further developing cyberinfrastructure components. Previous experience has shown that the development of a successful environmental observatory requires a close and early cooperation among domain scientists and computer scientists. The central role of science in this project is a key aspect of the CBEO prototype project.

It is critically important that any developments, abstractions, or applications envisioned by the computer scientists must be evaluated by domain scientists as contributory to the solution of pressing science questions. This requirement is in addition to the more readily recognized constraint that the needs/wants of the domain scientists be technically feasible and properly balanced against the time and effort that computer scientists must expend. Thus, parallel activities and active dialog are important.

CBEO Network and Cyberinfrastructure Components (CBEO:N)

The CBEO end-to-end cyberinfrastructure will deliver *observation data services* that include: (1) registering and annotating observation data resources of different types; (2) searching, accessing, and exploring the available metadata and data; (3) configuring and transforming the data into a common spatiotemporal framework; (4) using the transformed data as input to analysis, modeling, and visualization systems; and (5) providing personalized user interfaces to access the services. Such an end-to-end system would leverage several cyberinfrastructure components developed in other projects including the Geosciences Network, or GEON (Baru 2004; Zaslavsky 2006a,b), the National Virtual Observatory (Szalay et al. 2002), and the CUAHSI-HIS components (CUAHSI, 2007). Particularly the deployment of the CUAHSI-HIS time series data structure, or Observational Data Model, a database design specifically geared toward storing point time series data, together with a number of management tools, is of great importance to the development of the CBEO. The GEON cyberinfrastructure follows the principles of services-oriented architecture and represents a system of point-of-presence data and compute nodes, each with a respective stack of layered software components (GEON pack), for which a CBEO node will be deployed and extended.

Reusing core grid and data management services, including security and authentication management already implemented within the GridSphere portal, will let CBEO:N focus on its core contribution of providing support for additional types of data in CBEO, services for integrating disparate data sets and the means for interfacing these contributions with tools for analysis. Some specific contributions will include: (1) modeling and representing the various types of data being integrated into the CBEO (e.g., not only point-based time series, but also overflight satellite and model-generated data within volume elements as well as velocities and fluxes across interfacial areas); (2) developing web services that extend GEON registration and search functionality to data types associated with the many different types of observational and model-generated data that exist for the Bay; (3) defining standard interfaces for incorporating these additional data types in the software stack; (4) developing methods for transforming and packaging a user-configured fragment of virtual observatory as an input for analysis, modeling, and visualization software; (5) using powerful visualization software products, like the Integrated Data Viewer, IDV (UNIDATA 2008) or the FERRET system (PMEL 2006), that have been integrated with the GEON environment (SDSC 2006); and (6) defining communication interfaces between the CBEO services and applications, and the GridSphere portal framework used in GEON. The latter task is particularly important as a step toward making the emerging cyberinfrastructure components reuseable across different observation networks.

CBEO:E—Multicultural Student Development and K–12 Outreach

The educational and outreach component of the CBEO will be developed through activities that interface with existing education programs. The goal of this effort is to create new partnerships among scientists, science educators, teachers, and minority students to better convey information to a broad and diverse audience. More specifically, the CBEO prototype is being used to facilitate and assist three different education and outreach programs. First, the CBEO is working to facilitate data access and analysis for a Teacher Research Fellowship Program run through the Univ. of Maryland Center for Environmental Sciences—Maryland Sea Grant Environmental Science Education Partnership. This program involves teachers with scientists from partner institutions to enhance the teachers' understanding of science concepts and to develop classroom applications built on that work. Second, the CBEO is helping the development of new educational tools for use in the NSF sponsored Center for Ocean Science Excellence, Mid-Atlantic. This center integrates research and education by using hypoxia in estuaries to illustrate issues of ecosystem health. Finally, the CBEO is working directly with a long-standing NOAA sponsored program known as MAST (Multicultural Students At Sea Together), which is operated through Hampton Univ. (Cuker 2003). The MAST program involves a diverse crew of undergraduate and graduate students in a month-long cruise of the Chesapeake aboard a 16.2 meter (53 foot) sailing vessel as a means to combine the study of marine science, policy, the heritage of African Americans and Native Americans, and seamanship. Because MAST focuses on measuring water quality parameters such as oxygen and chlorophyll throughout the Bay, the link with the CBEO is again a natural one. Through the CBEO, MAST students will be able to examine their data in a broader context and with new tools, and also contributing new results to the overall project. By thus involving other educators and students in the development of the CBEO prototype, the project team hopes to gain additional insights into the particular system features that are most important for making the observatory useful not only for scientific research and resource management, but also for public education and outreach.

CBEO Test Bed Development (CBEO:T)

The CBEO:T (test bed) is a platform on which to test codes, synthesize data sets, develop interpolation routines, and also deploy and evaluate web services that operate within the larger observatory networks (CUAHSI and GEON). Locally the test bed consists of an Intel-based server that runs on the Microsoft Server 2003 operating system and has several software stacks installed, i.e., MS SQL Server 2005, ArcSDE Server, IIS, but also nonproprietary databases like postgreSQL. The test bed has more than 2.25 Tbytes of capacity and over 1 Tbyte of data loaded at this time.

The CBEO:T includes a database system that can be used to test the integration of model-derived data with observations across disparate databases. The system allows observational data to be queried against model data by interpolating the values of the point sources onto the model grid so they can be compared, correlated, and otherwise analyzed. This makes the test bed an ideal computational platform for hypoxia research that, when integrated with the GEON framework, makes it globally accessible to the scientific community.

The data sets of the CBEO:T are constructed from selected existing databases and data streams. These include archived output from Bay-scale water quality and hydrodynamic model runs; monitoring data from the Chesapeake Bay Program's Chesapeake Information Management System (USEPA 2007); MD DNR's "Eyes on the Bay" program shallow water data (MD DNR 2005); and aircraft remote sensing data (Harding et al. 1994). Specific tasks for this team include:

1. Integration of selected data sources and model data to resolve structural and semantic heterogeneity within the data, e.g., data format, units, spatial coordinate system, and ontologies (where appropriate).
2. Development of SQL language constructs for spatial joins that resolve spatial and temporal heterogeneity among point data, continuous data, and spatially averaged data.
   - Adaptation of the middleware SQL parser and optimizer developed for the NVO to translate data set joins to standard SQL operations and schedule these across multiple databases.
   - Building of a library of user-defined functions that implement interpolation functions for temporal and spatial joins executed at the databases, including the velocity-directed interpolation (III-C5).
3. Provision of web-services interfaces to the test-bed databases so that identified users may invoke, manage, and visualize scientific questions using the CBEO:T across the Internet.

Chesapeake Bay Science and Management (CBEO:S)

The CBEO:S team oversees the science connection to the planned cyberinfrastructure development. This science team will articulate and addresses well-posed science questions, first using the local test bed CBEO:T and then the CBEO network node CBEO:N. The team's main objective is to provide a "grounding" for the cyberinfrastructure developments, a task that previous experiences have shown to be crucial. With this approach, the domain scientists take the lead in the development of the CBEO rather than relying on cyberinfrastructure experts to dictate the creation of these types of information systems. Specific efforts are devoted to the following activities:

1. Examine the Hagy et al. (2004) finding (Fig. 3) in detail. Using the CBEO to explore anoxia development, using dynamic (velocity-directed) multivariate interpolation schemes to produce O_2 sources and sinks as variables. Explore the statistical significance of the perceived shift in loading: hypoxia relationships.
2. Perform a rigorous comparison of the 15 years of modeling output data and the integrated observational data set. This is in contrast to previous analyses that used only the fixed station data [Fig. 3(a)]. Focus on calibrations of source and sink terms as well as state variables.
3. More carefully compare the predicted versus observed hypoxic volume [Fig. 3(b)]. Use the integrated observational data set to seek reasons why the model fails to describe interannual variations. Examine miscalibration(s) under item 2 as potential cause(s) of problems.
4. Analyze the flux of organic matter and dissolved oxygen from the shallow water regions of the Bay using the integrated data set and the hydrodynamic model transport (velocity and diffusion) data.
5. Use the integrated data set to make improved estimates of primary production (Cerco and Noel 2004), a fundamental component of the models. Compare results with the integrated model-derived data.

Summary

The major goal of the CBEO project is to demonstrate that properly developed and deployed cyberinfrastructure can transform science by allowing researchers and managers to use data to explore questions in totally new ways. For example, it appears that many of the fundamental processes governing coastal hypoxia are not yet fully understood or adequately incorporated into current models for the Bay. The CBEO prototype aims to demonstrate that improved access and use of existing observational data, data streams, and modeling results can provide important new insights into these issues, and also facilitate future initiatives of research and education. We believe that this prototype is most effectively developed by an approach that allows environmental scientists and engineers to explore science and education questions at a "test-bed" level, and also supports parallel activities by computer scientists to develop the necessary cyberinfrastructure network for better shared use at regional and national scales. Such parallel development in connected activities allows each group to proceed without "waiting" for developments by the other but, more important, it also allows the development of interim results by both groups that can be used to inform the overall process. It is our hope that the approaches and tools developed in this prototype will also benefit a wide variety of other environmental research systems where improved integration of disparate kinds of measurements and modeling results are needed.

Acknowledgments

The CBEO project is funded by the National Science Foundation through Grant Number 0618986. The CBEO team would also like to thank Gary Schenk from the EPA Chesapeake Bay Research Program and Carl Cerco from the U.S. Army Corps of Engineers for making available large numerical modeling data sets as well as for their continued invaluable input to this project.

References

Baru, C. (2004). "The GEON grid software architecture." *Proc., 24th Annual ESRI Int. User Conf.*, Aug. 9–13, 2004, San Diego, ESRI, Redlands, Calif.

Bermudez, L., and Piasecki, M. (2006). "Community metadata profiles for the semantic web." *Geoinformatica J.*, 10 (March), 159–176.

Boesch, D. F., Brinsfield, R. B., and Magnien, R. E. (2001). "Chesapeake Bay eutrophication: Scientific understanding, ecosystem restoration, and challenges for agriculture." *J. Environ. Qual.*, 30(2), 303–320.

Cerco, C. F. (1995). "Simulation of long-term trends in Chesapeake Bay eutrophication." *J. Environ. Eng.*, 121(4), 298–310.

Cerco, C. F., and Cole, T. (1993). "3-dimensional eutrophication model of Chesapeake Bay." *J. Environ. Eng.*, 119(6), 1006–1025.

Cerco, C. F., and Noel, M. R. (2004). "Process-based primary production modeling in Chesapeake Bay." *Mar. Ecol.: Prog. Ser.*, 282, 45–58.

Chamberlin, D. (1998). *A complete guide to DB2 universal database*, Morgan Kaufmann, San Francisco.

Cloern, J. E. (2001). "Our evolving conceptual model of the coastal eutrophication problem." *Mar. Ecol.: Prog. Ser.*, 210, 223–253.

Collaborative Large-Scale Engineering Analysis Network for Environmental Research (CLEANER). (2007). "Water and environmental research systems network." ⟨www.watersnet.org⟩ (Apr. 2007).

Consortium of Universities for the Advancement of Hydrologic Sciences, Inc. (CUAHSI). (2007). "Hydrologic information system." ⟨www.cuahsi.org/his⟩ (Apr. 2007).

Cooper, S. R., and Brush, G. S. (1991). "Long-term history of Chesapeake Bay anoxia." *Science*, 254(5024), 992–996.
CRC. (2008). "Chesapeake community modeling program." ⟨wwww.chesapeake.or/ChesCommModel.html⟩ (July 2008), Chesapeake Research Consortium (CRC), Edgewater, Md.
Cronin, W. B., and Pritchard, E. W. (1975). "Additional statistics on the dimensions of the Chesapeake Bay and its tributaries: Cross-section widths and segment volumes per meter depth." *Special Rep. No. 42*, Chesapeake Bay Institute, The Johns Hopkins Univ., Ref. 75-3, Baltimore.
Cuker, B. E. (2003). "Minorities at sea together (MAST): A model interdisciplinary program for minority college students." *Current, J. Marine Education*, 18, 45–51.
Diaz, R. J. (2001). "Overview of hypoxia around the world." *J. Environ. Qual.*, 30(2), 275–281.
Ecological Metadata Language (EML). (2003). ⟨http://knb.ecoinformatics.org/software/eml/⟩ (July 2008).
Federal Geographic Data Committee (FGDC). (2007). "North American profile (NAP) of the ISO 19115 metadata standard." ⟨http://www.fgdc.gov/standards/projects/incits-l1-standards-projects/NAP-Metadata⟩ (Oct. 2007).
Hagy, J. D., Boynton, W. R., and Keefe, C. W. (2004). "Hypoxia in Chesapeake Bay, 1950–2001: Long-term change in relation to nutrient loading and river flow." *Estuaries*, 27(4), 634–658.
Harding, L. W., Itseire, E. C., and Esaias, W. E. (1994). "Estimates of phytoplankton biomass in the Chesapeake Bay from aircraft remote sensing of chlorophyll concentrations, 1989–92." *Remote Sens. Environ.*, 49(1), 41–56.
Harding, L. W., Magnuson, A., and Mallonee, M. E. (2005). "SeaWiFS retrievals of chlorophyll in Chesapeake Bay and the mid-Atlantic bight." *Estuarine Coastal Shelf Sci.*, 62(1–2), 75–94.
Howarth, R., et al. (2000). "Nutrient pollution of coastal rivers, bays, and seas." *Issues Ecol.*, 7, 1–15.
International Standards Organization (ISO). (2007). "Geographic metadata 19115:2003 Technical Committee 211." ⟨http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=26020⟩ (Oct. 2007).
Johnson, B. H., Kim, K. W., Heath, R. E., Hshieh, B. B., and Butler, H. L. (1993). "Validation of a three-dimensional hydrodynamic model of Chesapeake Bay." *J. Hydraul. Eng.*, 119(1), 2–20.
Kemp, W. M., et al. (2004) "Habitat requirements for submerged aquatic vegetation in Chesapeake Bay: Water quality, light regime, and physical-chemical factors." *Estuaries*, 27(3), 363–377.
Kemp, W. M., et al. (2005). "Eutrophication of Chesapeake Bay: Historical trends and ecological interactions." *Mar. Ecol.: Prog. Ser.*, 303, 1–29.
Kemp, W. M., Smith, E. M., Marvin-DiPasquale, M., and Boynton, W. R. (1997) "Organic carbon-balance and net ecosystem metabolism in Chesapeake Bay." *Mar. Ecol.: Prog. Ser.*, 150, 229–248.
LEADPortal. (2008). Linked environments for atmospheric discovery (LEAD), ⟨http://lead.ou.edu/⟩ (Jan. 2006).
Ludäscher, B., Lin, K., Jaeger-Frank, E., Bowers, S., Altinas, I., and Baru, C. (2006). "Enhancing the GEON cyberinfrastructure: Semantic registration, discovery, mediation, and workflows." *Internal Report of the Geosciences Network*, ⟨http://daks.ucdavis.edu/~ludaesch/Paper/geon-sms.pdf⟩ (July 2008).
Maidment, D. R., Zaslavsky, I., and Horsburgh, J. (2006). "Hydrologic data access using web services." *Southwest Hydrology*, 5(3), 16–17.
Malik, T., Burns, R., and Chaudary, A. (2005). "Bypass caching: Making scientific databases good network citizens." *Proc., Int. Conf. on Data Engineering*.
Malik, T., Szalay, A. S., Budavri, A. S., and Thakar, A. R. (2003). "Sky Query. A webservice approach to federate databases." *Proc., Conf. on Innovative Data Systems Research.*
Malone, T. C., Boynton, W., Horton, T., and Stevenson, C. (1993). "Nutrient loading to surface waters: Chesapeake case study." *Keeping pace with science and engineering*, M. F. Uman, ed., National Academy Press, Washington, D.C., 8–38.
Malone, T. C., Crocker, L. H., Pike, S. E., and Wendler, B. W. (1988). "Influences of river flow on the dynamics of phytoplankton production in a partially stratified estuary." *Mar. Ecol.: Prog. Ser.*, 48, 235–249.
Maryland Dept. of Natural Resources (MD DNR). (2005). "Eyes on the bay." ⟨www.eyesonthebay.net⟩ (Jan. 15, 2006).
myGEON. (2006). ⟨http://www.geongrid.org/mygeon.html⟩ (Jan. 2006).
National Science Foundation (NSF). (2005). "Cyberinfrastructure of environmental observatories: Prototype systems to address cross-cutting needs (CEO:P)." ⟨http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=13517⟩.
NEON. (2007). "Transforming ecological research, and forecasting ecological change." National Ecological Observatory Network, ⟨http://www.neoninc.org⟩ (Sept. 2007).
Nixon, S. W. (1995). "Coastal marine eutrophication: A definition, social causes, and future concerns." *Ophelia. Int. J. Marine Biol.*, 41, 199–219.
OCEAN.us. (2005). "Data management and communications plan (DMAC) for research and operational integrated ocean observing systems (IOOS)." ⟨http://dmac.ocean.us/dacsc/imp_plan.jsp⟩.
OOI. (2008). *Oceans observatories initiative (OOI)*, ⟨http://www.oceanleadership.org/ocean_observing⟩ (July 2008).
Pacific Marine Environmental Laboratory (PMEL). (2005). "FERRET: An interactive computer visualization and analysis environment." ⟨http://ferret.pmel.noaa.gov/Ferret/⟩ (Jan. 2006).
Rabalais, N. N., Turner, R. E., and Scavia, D. (2002). "Beyond science into policy: Gulf of Mexico hypoxia and the Mississippi River." *BioScience*, 52(2), 129–142.
Roman, M., Zhang, X., McGilliard, C., and Boicourt, W. (2005). "Seasonal and annual variability in the spatial patterns of plankton biomass in Chesapeake Bay." *Limnol. Oceanogr.*, 50(2), 480–492.
Rosenberg, R. (1990). "Negative oxygen trends in Swedish coastal bottom waters." *Mar. Pollution Bull.*, 21(7), 335–339.
SDSC. (2006). "The Geosciences network, GEON: Building cyberinfrastructure for the GeoSciences." ⟨http://www.geongrid.org/⟩ (Jan. 2006).
Szalay, A. S., et al. (2002). "The SDSS Skyserver: Public access to the Sloan Digital Sky Server data." *Proc., ACM SIGMOD Int. Conf. on Management of Data*, Association for Computing Machinery's Special Interest Group on Management of Data.
Ueng, S.-K, and Wang, S.-C. (2005). "Interpolation and visualization for advected scalar fields." *Visualization*, IEEE, Piscataway, N.J.
UNIDATA. (2008). "Integrated data viewer (IDV)." ⟨http://www.unidata.ucar.edu/software/metapps/⟩ (July 2008).
USEPA. (2007). "Chesapeake Bay program CIMS." ⟨http://www.chesapeakebay.net⟩ (Nov. 26, 2007).
Yang, Q., and Parvin, B. (2003). "High-resolution reconstruction of sparse data from dense low-resolution spatio-temporal data." *IEEE Trans. Image Process.*, 12(6), 671–677.
Zaslavsky, I., et al. (2005a). "Grid-enabled mediation services for geospatial information." *Next generation geospatial information. From digital image analysis to spatiotemporal databases*, P. Agouris and A. Croitoru, eds., Taylor & Francis, London, 15–24.
Zaslavsky, I., Memon, A., and Memon, G. (2005b). "Integration across heterogeneous spatial data and applications within a large cyberinfrastructure project." *GIS Devel.*, 9(5), 28–31.
Zhang, Y., and Akambhamettu, C. (2003). "On 3-D scene flow and structure recovery from multiview image sequences." *IEEE Trans. Syst., Man, Cybern., Part B: Cybern.*, 33(4), 592–606.