Information infrastructure for integrated ecohydraulic and water resources modeling and assessment

R. Wallace, K. Pathak, M. Fife, N. L. Jones, J. P. Holland, D. Stuart, J. Harris, C. Butler and D. R. Richards

ABSTRACT

Watershed management increasingly requires ecohydraulic modeling and assessment within a regional context, rather than on a project-by-project basis. Such holistic modeling and assessment require evaluation capabilities across multiple temporal and spatial scales. Thus, modeling and assessment tools must be integrated in a scientifically and computationally effective infrastructure. The US Army Engineer Research and Development Center, in concert with the Hydrologic Engineering Center and its academic partners, including Brigham Young University, is establishing a comprehensive set of hydroinformatics modeling and assessment tools for ecohydraulic and water resources management applications, all linked based on a common data and information infrastructure. This paper presents the attributes of this information infrastructure and compares it with the analogous integration initiatives elsewhere.

Key words | computational infrastructure, computer-based modeling and assessment tools, cyberinfrastructure, ecohydraulics, environmental hydroinformatics

R. Wallace
Coastal and Hydraulics Laboratory,
US Army Engineer Research and Development
 Center,
3909 Halls Ferry Road, Vicksburg, MS 39180-6199,
USA

K. Pathak
J. P. Holland (corresponding author)
D. Stuart
C. Butler
D. R. Richards
Information Technology Laboratory,
US Army Engineer Research and Development
 Center,
3909 Halls Ferry Road, Vicksburg, MS 39180-6199,
USA
Tel:+1 601 634 2644
E-mail: *Jeffery.holland@us.army.mil*

M. Fife
N. L. Jones
Environmental Modeling Research Laboratory,
242 Clyde Bldg.,Brigham Young University,
Provo, UT 84602, USA

J. Harris
Hydrologic Engineering Center,
US Army Engineer Institute for Water Resources,
609 2nd St, Davis, CA 95616-4687, USA

ABBREVIATIONS

BYU	Brigham Young University
CDF	Common Delivery Framework
CplComp	The Coupler Component of ESMF
DES	Description file
DIAS	Dynamic Information Architecture System
DSS	Data Storage System
EPA	U.S. Environmental Protection Agency
ERDC	U.S. Army Engineer Research and Development Center
ESMF	Earth Systems Modeling Framework
ESRI	Environmental Systems Research Institute, Inc.
FRAMES	Framework for Risk Analysis in Multimedia Environmental Systems
GID	Global Input Data
GIS	Geographical Information System
GridComp	The Gridded Component of ESMF
GUI	Graphical user interface
HEC	U.S. Army Engineer Institute for Water Resources Hydrologic Engineering Center
HPC	High performance computing
MIMS	Multimedia Integrated Modeling System
MMS	Modular Modeling System
OpenMI	Open Modelling Interface and Environment
SDO	Service Data Objects, Oracle's object for storing geometric data
SMS	Surface water Modeling System
TIN	Triangular Irregular Network
USACE	U.S. Army Corps of Engineers

WAT   Watershed Analysis Tool
WMS   Watershed Modeling System
XMDF   Extensible Model Data Format

INTRODUCTION

The US Army Corps of Engineers (USACE), through its Engineer Research and Development Center (ERDC), Hydrologic Engineering Center (HEC), and academic partners such as Brigham Young University (BYU), is developing a suite of modeling and assessment tools to aid decision making on a systems (watershed, river basin, coastal zone) level. Ecohydraulic challenges such as the Florida Everglades, the Upper Mississippi River, the Columbia/Snake River System, and the Missouri River require watershed and basinwide systems analysis where tradeoffs across multiple spatial and temporal scales can be evaluated. These problems are multidisciplinary in nature, necessitating the integration of models and assessment tools (hereafter referred to interchangably in this paper as "models") of highly differing fidelity, dimensionality, and scientific and computational complexity. As an example, the environmental assessment of the Upper Mississippi River requires single- and multi-dimensional hydrodynamic and hydraulic models, ecological models for multiple species (each of differing dimensionality), habitat suitability models, and physical-forcing models associated with navigation effects – each linked in some fashion with each other, with geospatial databases, and with operational information systems.

MODEL INTEGRATION ISSUES

Any computational modeling and assessment system must address certain issues in order to integrate disparate components: method of exchanging data between connected models, tools and databases, means of setting up an integrated model, modes of control when the integrated model is being executed, decisions on whether individual models run sequentially or simultaneously, the operating systems to be used to execute the models on a given computational platform, and how the results are visualized. These issues will be addressed in the following sections.

Data exchange

Models developed for differing applications vary widely in terms of format and data requirements. As a consequence, data exchange can be challenging. Data are typically exchanged using files or databases where the output from one model becomes the input for another model. In such a case it is necessary to ensure that all input parameters are provided in the output file and that the data are valid and properly formatted.

Data can also be passed directly from model to model during execution using a linking object to connect parameters from each model in the system. Sophisticated linking objects may rely on software connectivity protocols such as .NET, Java, or CORBA, but they can also use a simple data file approach. Regardless of the data transfer mechanisms, execution time coupling usually requires that models be linked at a specific time at a specific location. This often requires that interpolation and extrapolation functions be employed so that models send and receive information at appropriate temporal and spatial resolutions.

Model setup

Modern modeling and assessment systems provide utilities to assist the user in setting up the integrated model. Fully automated linkage can occur only when the models/tools that are linked within the system conform to a standard. Systems can either require models to use a specific file format or require models to be modified to meet their standard. Most integrated systems modify existing assessment capabilities by wrapping them with a pre- and post-processor. The pre-processor modifies input to the model so that it matches a required input format, while the post-processor will change the computed output to conform to the integrated system. The advantage of defining a wrapper is that the analysis executable remains unchanged.

In addition to the requirements involved in linking computational tools, there must also be some way for the user to specify how the models are to be linked. This is usually done through a graphical user interface (GUI) that allows the user to see a visual representation of linkages and dependencies between system components.

Control of model execution

Once models have been linked in an integrated system, the system must know how to pass control between the various models and when to pass control back to the user. This will depend primarily on how the data exchange has been organized and whether the models run sequentially or concurrently. A system that runs models sequentially is much simpler than one that exchanges data between models as they are run concurrently. While this former system may be less complicated, it also does not allow temporally dependent optimization (feedback) to be performed.

Operating systems

The operating environment utilized by each computational model is another factor influencing the development of an integrated system. Models that run on a PC or workstation-only environment may not be suitable for analyzing a large-scale basinwide problem at very high temporal and spatial resolutions. In these situations it is important to know if the simulation can be run in a high performance computing (HPC) environment or if the models can be connected across distributed networks. The more flexible the model linking system is, the greater the ability the user will have to create more extensive integrated models.

Model integration types

The basis for a modeling framework is the data communication that allows each model to communicate with other models. Two primary forms of data communication can take place when a model integration system is being developed: linking and coupling. Linking adopts a pass-through data communication mechanism. In this scenario, Model A completes an analysis based on input conditions and then saves the model results in some type of output data structure. Model B then reads as input the output data from Model A and executes its analysis. Information is not passed between Model B and Model A in this scenario.

Model coupling provides a more integrated approach that may be necessary when feed-back from Model B is required to properly compute an analysis in Model A. In this approach Model A and Model B execute simultaneously and data are passed between the models after completion of specified analysis intervals. This type of model integration requires that the models can communicate during run-time. Various methodologies have been adopted to allow this type of communication including data files, CORBA, .NET and Java. Model coupling requires a substantial amount of work to standardize the methods by which data is passed between the models to ensure rapid and accurate data transfer between models.

Model execution

Two basic methodologies are employed for controlling a combined modeling scenario: master control and object control. In master control, the entire scenario is orchestrated by a master controlling algorithm that initiates when the first and subsequent computational models will run, identifies where intermediate and final data is stored, and initiates contact with the user when model execution is complete. This type of execution methodology requires that the master controlling algorithm has a complete understanding of all parameters of the scenario and can fully adapt the execution logic depending on modeling results.

The other primary type of model execution methodology acts more like an object-oriented program. In this connection scenario, execution is initiated by a "trigger" program and each subsequent model begins execution when data is required from the other computational program. Model control is passed to the called program, which executes until it has satisfied the data request of the calling program. Model control and the data then return to the calling program. This type of execution methodology provides a complete abstraction between the model requesting data and the model providing the data. Furthermore, the requesting executable does not need to know what type of program or data source is providing the data, as long as data is returned in a format and at the location that was specified. Spatial and temporal interpolation is controlled by the connection between the models, not necessarily by the models themselves.

EXISTING INTEGRATED MODELING SYSTEMS

Integrated systems development for ecohydraulic modeling and water resources management is an active area of

research. This section provides an overview of some of the more well-known model-linkage systems.

FRAMES

The Framework for Risk Analysis in Multimedia Environmental Systems (FRAMES) (Whelan *et al.* 1998) was developed by the Pacific Northwest National Laboratory to integrate models across scientific disciplines to help assess environmental, safety, and health risks. FRAMES exchanges data through input and output files. Modules are linked in a process flow diagram and executed in a sequential fashion. The output file of one module is used as input for the next module in the sequence.

To be certain that the information exchanged is valid, each module has a Global Input Data (GID) file that contains the connection and user input data for the module and a Description (DES) file, which describes the module so it can be properly linked. The DES file contains a description that can be seen by users as they select a module and includes information such as what the module is typically used for, what the limiting assumptions are, length and time scales, references to underlying formulations, verification and validation documents (if they exist), hardware requirements, and contact information. A module can be included in a simulation either by creating it in FRAMES or by wrapping a model, wrapping the corresponding user interface of the model, and creating a DES file so that FRAMES will recognize it.

The FRAMES user interface is organized into two workspaces, as shown in Figure 1. In the lower workspace, modules are linked to form simulations; in the upper workspace, available data for multiple modules (such as a database of chemicals) is represented. These data are shown separately because they may not need to be linked to a module individually. Linked models are created by dragging modules from a selection tree into the workspace. Modules can be moved, deleted, connected, and disconnected with the mouse. Modules that are added to the workspace are given a status light to show the completion status of the module. A red light means the module has been included in the integrated system. A yellow light indicates the input data are complete,
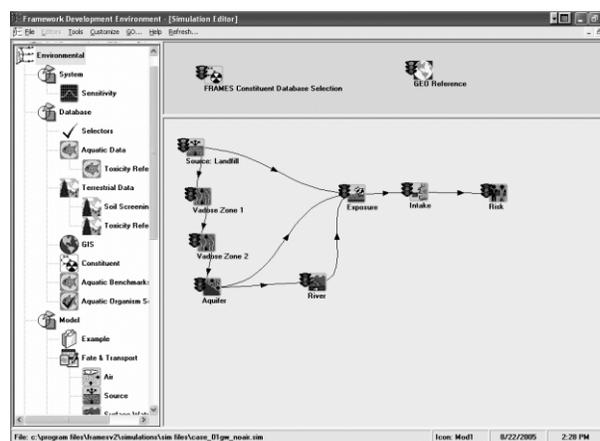

Figure 1 │ FRAMES user interface.

while a green light means the code calculations have been successfully run. Each module in FRAMES has its own user interface. Data can be input to the module by bringing up its user interface.

Once all the data have been input for a module, it can be executed. Each module can be run individually or the entire sequence can be run together, as long as the input for each module has been supplied. After making changes to the module input, the entire sequence or a subset of the models can be re-executed. It is not possible to do cyclic or concurrent processes in FRAMES since each new module depends on the results of the preceding one. After a module has been executed, its results can be reviewed by right-clicking on the module. A set of available viewers for each of the module output types are listed. A text view and at least one graphical view are always included. The FRAMES User Interface and all other user-defined source modules operate on Windows 95 or higher.

DIAS

The Dynamic Information Architecture System (DIAS) (Campbell & Hummel 1998) was developed by the Argonne National Laboratory Decision and Information Sciences Division. This product was designed to address the complexity and dynamics of real-world systems. It brings together models and applications from multiple disciplines.

In order to make DIAS a flexible system, data are exchanged separately from the models. An integrated model

system consists of several domain objects that represent real-world entities and models or applications that express the behavior of those objects. The domain objects contain an abstract description of the behavior, but all implementation details are contained in the models and the applications. The models do not exchange data directly with each other, but are linked to the domain objects through a process object.

Process objects are the only objects that contain information about both domain objects and about their implementation. They package/unpackage data that are passed to/from the models, perform disaggregation/aggregation of data, transform coordinate systems, and do unit conversions.

External models can be used in DIAS by defining the application. This is done by specifying parameters for the domain object, selecting processes within the domain object, and establishing input/output parameters for these processes. Then a model source code must be created along with a model controller or wrapper. Domain objects must also be built or edited to correspond with the model. In addition, a process object must be built for each process that is represented in the source code.

An integrated model is created by implementing a conceptual frame containing all the entities that will be needed in the process. The domain objects can be instantiated with data from external sources. Models and applications can then be chosen by the user, or the Context Manager can automatically select models, applications, and databases that are already integrated into DIAS. Figure 2 is

a representation of how all the objects are linked together in the DIAS analysis frame.

The Discrete Event Simulation Manager processes events in a time-ordered sequence. Each event has a time stamp associated with it and may contain data. Events are created by the behavior of the domain object and by user interaction. Each process is run according to the time stamp of the event. Models in DIAS can be distributed across a network of processors or can be run on a parallel distributed memory computer.

MIMS

The Multimedia Integrated Modeling System (MIMS) (Fine *et al.* 2002) developed by the US Environmental Protection Agency (EPA), was created to help link models created by different agencies within the EPA, including models from different disciplines.

Since MIMS depends on DIAS, much of the data exchange process is the same. Relationships between models are defined by a connection. The sequence of these connections determines the order of operations in the simulation. Parameters are passed between models through these connections as well. A module assumes the same value for a specified parameter as the domain object it is connected to. Tools are also available to make parameter connections
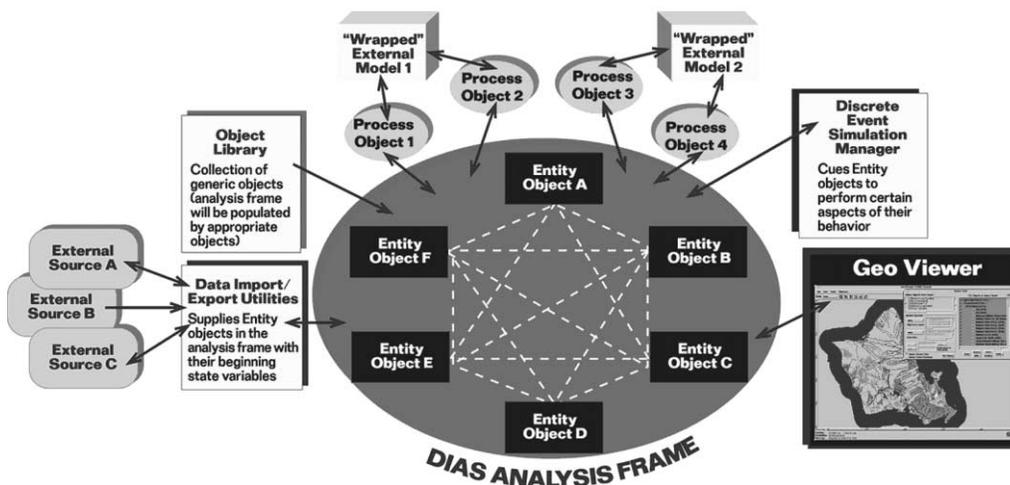

Figure 2 | DIAS object interactions.

explicitly. Each module has a control file to keep track of the parameters and their connections.

The MIMS framework layers on top of the DIAS user interfaces to facilitate bringing existing models into the simulation. A module is prepared for linking by creating a control file and defining the parameters that will be involved in the simulation. After the module has been defined, its corresponding domain object can also be created. Processes that will occur in the simulation must be defined with the domain object.

To create a linked model, domain objects are selected and flow arrows are created to indicate how information is exchanged between modules. Parameters can be explicitly connected at this time. When all inputs are supplied for each domain object, the model can be run. Figure 3 demonstrates how an integrated model is created in MIMS. The tree lists available domain objects and their corresponding processes. At the bottom is the parameter table, which links parameters to domain objects. This table contains other pertinent information about the parameters

as well. The order of execution is defined by the connections established between modules. The entire integrated model is executed before control returns to the user. To view results, an analysis domain object must be included in the linked model. Just as for the individual models, the type of analysis that needs to be done must be specified. The MIMS Framework is designed to operate in the Java Runtime Environment (JRE) Version 1.4. It can be used on both Windows and UNIX systems.

MMS

The US Geological Survey's National Research Program has developed the Modular Modeling System (MMS) (Leavesley *et al.* 1996) to create the ideal model for a specific application by selecting the most appropriate algorithms from a set of available functions.

Each module stores parameters, variables, and dimensions in three separate internal databases, which can be
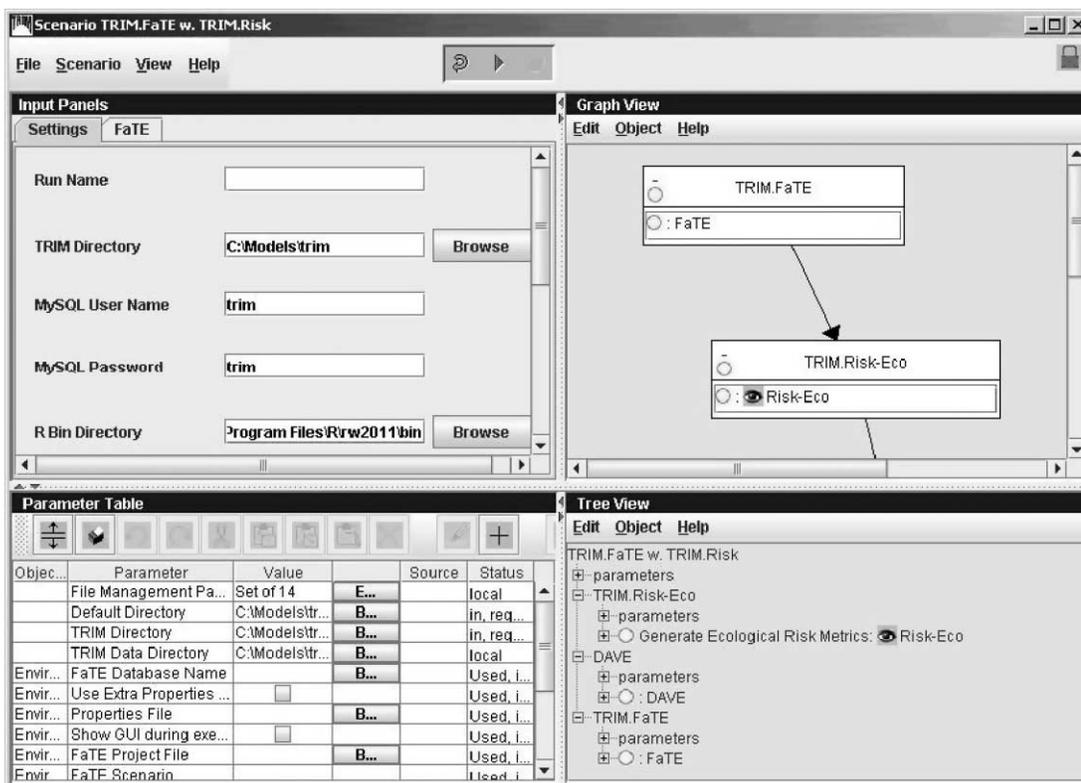

Figure 3 | MIMS user interface.

passed between modules by library functions. The integrated model also has a control database that saves the status of all features so that it can be reopened with all of its parameters defined. Additionally, all models have an environment file that specifies directory paths, system files, and other MMS related utilities. Most importantly, this file specifies the location of the control file or database.

A user directory must first be created so that modules can be added to it. FORTRAN or C modules can be chosen from the module library and linked together using an interactive model builder interface. New modules can also be created in MMS, and existing code can be modified to be used as a module in MMS. When a model is first initiated, all the modules are read through to create the parameters database. Functionality is provided to edit parameter files and data files.

Multiple files can be read simultaneously for a single model run. Data lines in all input files must be ordered from earliest to latest. When multiple files are used, the first line from each file is read and the earliest time stamp is selected. When data are needed for the next time-step, the next line in the file is compared with the other files to select the earliest time stamp. This process continues until there are no more data to be processed.

MMS provides several graphs to view the results of the model run. Sensitivity analysis graphs allow the user to view contours and plots as well as a simple three-dimensional (3D) view. Probability distribution curves can also be viewed for the maximum flow, volume, time to reach the maximum value, and time to fall below a user-defined threshold.

MMS was originally developed for UNIX-based workstations and uses X-windows and OSF-Motif for the GUI. Operating systems that have been successfully used are Sun (both SunOS 4 and Solaris), Data General (DGUX), Hewlett-Packard (HP-UX), IBM (AIX), Silicon Graphics, and PC (Linux). More recently, MMS has been precompiled to run under the Windows operating system using the CYGWIN software package.

OpenMI

The Open Modelling Interface and Environment (OpenMI) (Blind & Gregersen 2005) was developed by the European HarmonIT project to provide a framework flexible and generic enough so that any model can easily be linked to another.

In OpenMI, models, databases, and tools exchange information during run time. This is achieved with a component-based design, where all components have the same minimum properties, methods, and events. A model needing information requests a value set from another model or database for a specific time at a specific location. Information is passed directly between models without the use of intermediary files. Because of this direct data exchange, interpolation and extrapolation functions must be provided to calculate the necessary data.

To be used in an integrated system, models must comply with OpenMI Standards. Existing models can be wrapped to fit these standards. In order to prepare a set of linkable components with the associated links, OpenMI has prepared an editor that allows the user to select all the linkable components, create the links, configure the links, and set any other parameters for the simulation. Linkable components are stored and can be easily accessed. The link is formed simply by drawing a connecting line from one model to another with the mouse. Double-clicking on the link brings up a dialog that allows the user to configure the link by defining the information each module needs and what results each module will return. An XML file provides metadata for each linkable component to allow modelers to know how to link models together.

While models may have their own unique distribution of time-steps, a model in OpenMI should be able to return data at any time or location. The data exchange process is completely automated, and control does not return to the user until the model run is complete. Models run simultaneously and share information directly. OpenMI currently relies on the .NET framework and is limited to computers running the MS Windows operating system. A Java implementation is currently (2006) under development.

ESMF

Another methodology for developing integrated model scenarios is the Earth Systems Modeling Framework (ESMF) (Collins *et al.* 2005). This system is being developed

by a multiagency working group with additional funding from the National Science Foundation (NSF). The ESMF defines an architecture for composing complex, coupled systems and includes data structures and utilities for developing individual models. ESMF is dedicated primarily to developing regional and global scale integrated computational models. ESMF defines the protocols and connection methodologies that allow separate computational modules to operate as a single executable. ESMF is especially well suited for development of systems that operate in an HPC environment.

ESMF works on the idea of components and offers two: the Gridded Component (GridComp) and the Coupler Component (CplComp). The GridComp represents data or the physical domain being modeled, and the CplComp transfers and transforms data between two or more GridComp components. Physical processes are also implemented as components and model developers incorporate these components at the library level by making calls to the ESMF library functions. File I/O is not addressed in ESMF and is the responsibility of the model developer.

Because ESMF is developed at the library level, the completed system operates as a single executable. ESMF is developed primarily in FORTRAN but can be accessed by other programming languages as well. It currently is supported on most Unix and Linux platforms.

ERDC COMPUTATIONAL INFRASTRUCTURE

The ERDC computational infrastructure in support of ecohydraulic modelling and water resources management and decision making is shown conceptually in Figure 4. The major levels of this infrastructure are as follows:

- Integrated framework: a common information technology structure that can deliver information as well as modeling and assessment tools.
- Data handling, fusion, aggregation, management, and mining: key technologies that enable users to organize, analyze, present, and reuse data across multiple application areas.
- Modeling and assessment (M&A): computational tools that model/assess key components of physical or empirical processes.
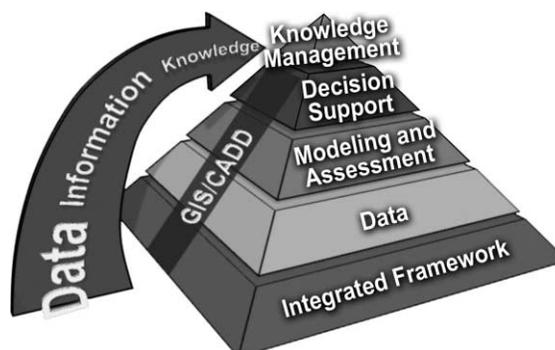

Figure 4 │ Levels of the ERDC water resources/ecohydraulic computational infrastructure.

- Decision support: set of capabilities that synthesize and present data and information in a variety of manners that directly support the decision-making process, provide different courses of action, or analyze alternatives.
- Knowledge management: the information infrastructure that provides a collaborative, information-rich environment among multiple stakeholders.

The following subsections discuss the framework, data, and modeling/assessment levels of the ERDC computational infrastructure. Information regarding decision support and knowledge management is left for a future paper.

Framework level

The framework efforts described in this paper are focused on creating technologies that allow selected models to be delivered and accessed in the way most appropriate for the model. It also allows for model-to-model communication using remote invocation and data sharing through standardized formats. The technologies that are being created are divided into four major categories:

- Geographical information system (GIS)
- Web-based network
- One-dimensional (1D) analysis tools
- Multidimensional (MD) analysis tools

GIS environment

A large portion of the models in this tool suite rely on some type of geospatial information. To support this need, a framework for storing, retrieving, indexing, and searching

GIS data is being developed. This framework is based primarily on Environmental Systems Research Institute, Inc. (ESRI), Arc 9 platform. The framework encompasses both desktop and server-based capabilities. The desktop portion gives guidance on creating ArcGIS (http://www. esri.com/software/arcgis/) desktop extensions that give access to remote data services such as those in the DataNET (covered in the Data Management subsection below). The server portion called the Data Repository (also covered in the Data Management subsection below) supplies a framework for storing large data files in their native format with a relational database for storing metadata about each file. The database is used to index and query the data sets. Oracle Service Data Objects (SDO – Oracle's object for storing geometric data; http://www.orafaq.com/faqsdo.htm) and ESRI ArcSDE (server software product used to access massive multiuser geographic databases stored in a relational database; http://www.esri.com/software/arcgis/ arcsde/) are used to create tools for visualizing and downloading the data files as needed.

Web-based networks

In many instances, modeling and assessment results need to be made accessible remotely, either through an interactive Web site, or through programmatic remote calls. To meet these needs, a Web Framework is being created that helps developers to accomplish both tasks. For programmatic access to models the W3C standard of Web Services (http://www.webservices.org) has been adopted. This allows remote, cross-platform calls to be made between models and allows remote access to data sources on an as-needed basis. Web services use standard Web-based protocols that can retrieve data from any hardware/software platform.

As part of a previous ERDC effort, called the Common Delivery Framework (CDF) (www.erdc.usace. army.mil/pls/erdcpub/www_welcome.navigation_page?tm p_next_page=61035), the idea of using Web services as part of a modeling infrastructure was studied, and several services were created that are being used presently. The first is an authentication service that will allow model authors to control access through one of several "authentication sources", including Microsoft's Active Directory for internal

Corps of Engineers users, Army Knowledge Online for Army users, and a catchall local user system for external users. The second is a document control service called Content Services. Content Services allows for programmatic publishing and retrieval of unstructured data including metadata. These services incorporate the authentication service to control access to different areas of the taxonomy. An interactive publishing application is also available for use.

For interactive models, a portal framework has been created based on the open-source portal Rainbow (http:// www.rainbowportal.net). Modules have been added to the portal to give access to some in-house services such as authentication and content services, and access to scientific data through DataNET Web services.

In addition to these capabilities, the Web framework supplies access to U.S. Department of Defense HPC resources through a program called ezHPC (https:// solutionhpc.erdc.hpc.mil). ezHPC is available both as a desktop and a Web-based application. It is designed to give people not comfortable with *nix style command line models easy access to these systems. Using a GUI-driven process, ezHPC allows the user to upload data, set up model runs, monitor batch jobs, and view results right from the desktop workstation.

1D analysis tool

The 1D space is served by the Watershed Analysis Tool (WAT). This application is Java-based, using client/server architecture to connect inputs and outputs of the various models. The user is allowed to select reaches of waterways to apply a model to, and to create nodes that designate transition from one model to the next. All I/O from the models is directed to a server where results are stored and made available for the next model. These results are persisted, allowing them to be reused through multiple runs as well as shared between multiple users.

MD analysis tools

The baseline software being used for the multidimensional framework is the XMS suite of software. XMS is a generic term that encompasses the following software:

- The Groundwater Modeling System (GMS) (http://chl. erdc.usace.army.mil/Media/3/7/2/GMSFactSheet.pdf)
- The Surface water Modeling System (SMS) (http://chl.erdc. usace.army.mil/Media/3/9/7/SMS8-Fact 20Sheet.pdf)
- The Watershed Modeling System (WMS) (http://chl.erdc. usace.army.mil/Media/4/0/4/WMS-Fact 20Sheet.pdf).

These software packages represent over a decade of research and development carried out at ERDC. The XMS series is written for Microsoft Windows in Microsoft Foundation Code (MFC) C. The base capabilities are being taken out of these applications, and they are being transformed into modular, cross-platform libraries that can be used in any supported model to create accurate, consistent results in a standardized format. These results can then be shared and archived.

There are two primary efforts to achieve these goals. The first effort is to create libraries using standard ANSI C++ with a complete build system that supports all the major computing platforms (Windows, *nix, OSX, HPC, etc.). The initial library will include an application programming interface for XMS mesh objects and will include all the basic methods/properties required for editing and manipulating 1-D, 2-D, and 3-D meshes including creating meshes, adding nodes and elements, modifying or deleting nodes and elements, traversing node and element lists, spatial queries, boundary searches, and file I/O.

The second effort targets file I/O and uses a generic modeling data format Extensible Model Data Format (XMDF) (Jones *et al.* 2004) as the format for storing and manipulating scientific data sets such as grids, meshes, and Triangular Irregular Network (TINS). XMDF is a cross-platform (hardware) binary file based on the Hierarchical Data Format 5 (HDF5) (http://hdf.ncsa.uiuc.edu/HDF5). It supports hierarchal storage of data, metadata (including georeferencing), and built-in compression and is already used extensively in the XMS suite. By transitioning other models supported to use it for I/O, XMDF will make model integration/communication much easier.

Data level

Data management is critical to support any integrated tool suite for systemic water resources and ecohydraulic

management. Managing and using data is considered as one of the most complex, tedious, and time-consuming efforts associated with ecohydraulic planning and management. Data are used by all applications, available in various formats, and consumed by the applications in various formats. Sometimes an application needs data from various sources (in different formats) to feed into a single model. Fresh, complete data are imperative to run the applications and models successfully. The data are generally acquired by ftp, http, or actual physical media transfer leading to inefficient and inconsistent usage of data and data sources. Furthermore, it is cost prohibitive for all the users and applications to collect and maintain the freshness of the data independently, leading to varying and conflicting results from the same models run by different sources.

Five types of data have to be accessed, integrated and provisioned in this modeling suite:

- Geospatial/GIS data (Oracle SDO, ESRI ArcSDE, etc.)
- Scientific/Model data stored in XMDF format
- Scientific/Model data not stored in XMDF
- Unstructured data (.doc,.jpg,.pdf, etc.)
- Relational Database data (Sediment, Chemistry, etc.)

A two-pronged approach is being used in addressing the complexities of data retrieval, processing, assimilation, provisioning, and usage. The Data Repository retrieves data from specific data sources and holds it in a centralized database in the native format. The DataNET employs a network-centric approach to streamline and standardize the acquisition, dissemination, and management of data across all USACE business areas by using Web services and a customized ArcGIS toolbar extension. Though these are two different mechanisms being used for data manipulation, the interface is seamless and transparent to the user and applications (with single sign-on). The Data Repository and DataNET can be accessed through https://swwrp.usace.army.mil.

Data repository

Data from sources that do not provide a reliable format or delivery mechanism must be held in a local database for programmatic on-the-fly access by users and applications. Sometimes, the data come in an unstructured format (such as Microsoft Excel, text, or PowerPoint files) that cannot be

directly ingested by users and applications. To resolve these nonstandard formats and delivery of data, the Data Repository morphs and pre-processes the source data into a standard format that is user- and application-friendly. At regular established intervals (negotiated with the individual data sources) the Repository either pulls out the fresh data or pushes back the fresh data that it morphs into a standardized format using consistent algorithms. Although the Data Repository is the low-hanging fruit for data assimilation, it remains a resource intensive process (both machine and human) to access, integrate, and supply the data. The evolutionary product to the Data Repository is the DataNET (described next), which provides a more adaptive, modular and sustainable solution to data management.

DataNet

The DataNet acts as a central clearinghouse portal for data acquisition for USACE, funneling data from other US Federal and state agencies (U.S. Geological Survey, National Aeronautics and Space Administration, EPA), and the industry (ESRI, Microsoft) without replicating the data in the Data Repository, thereby significantly reducing sustainment costs. The primary objectives of the DataNet are to:

- Develop, promote, and deploy a common net-centric framework that provides a consistent interface to data sources internal and external to USACE
- Formalize connectivity to USACE data sharing partners (Federal and State governments, universities, industries, and the public)

The DataNet employs the same Web services described previously in the Frameworks subsection to retrieve and provide data from various data sources without caching the data locally. The ERDC team develops Web services for existing and new data sources and makes them available via two vehicles of delivery (described below).

The DataNet comes in two flavors; the *Data Browser* and an *ArcGIS toolbar* extension. Figure 5 shows the Data Browser as a thin-client Web application developed in the .NET architecture. The user accesses the application over an https secure connection and is able to pull up point-data that shows the locations of all the data available for download. This is accomplished by storing metadata on the various data sources in a local database, which is

refreshed periodically to ensure freshness of data. The metadata contains such information as owner, creation data, survey bounds (geo, polygon, etc.), collector, file type, and original and encoded filenames. The metadata are displayed using a map viewer and can also be searched using strings. The map viewer has standard ESRI functionalities of zooming in and out to provide a more granular view of the visual depiction of data. Once the user has decided on what data is needed, the option for downloading the data can be clicked, in which case the user is redirected to the original source of the data in a seamless process.

Figure 6 shows the ArcGIS toolbar extension called Web Service Toolbar operating in the ESRI ArcInfo environment. Once installed, the toolbar provides the user with a suite of data sources that can be accessed programmatically, the data sources appearing as standard widgets as shown, with standard tool tips. When a data source is clicked, a Web service is called and the data from the original data source is pulled into the ESRI environment, ready to be used by the user.

By employing the two mechanisms of data retrieval, integration and provision using the Data Repository and the DataNet, science and engineering client applications such as computational models, Web applications, GIS applications, and portable devices can directly access the freshest heterogeneous data from various data sources (without local caching) to support modeling and assessment.

Modeling and assessment level

USACE has invested in a significant number of computational models and assessment tools. Many of these were developed to solve a specific water resources problem. However, the USACE does not have sufficient funding to support every ecohydraulic, water resources, environmental, or sedimentation computational model that has been developed. In response to this funding reality, and in an effort to establish a baseline of computational models that will be used to solve large domain or basinwide problems, a set of models has been established that are supported by the modeling and assessment suite discussed in this paper. The list is intended to reflect both the existing state of practice within the USACE offices and the emerging
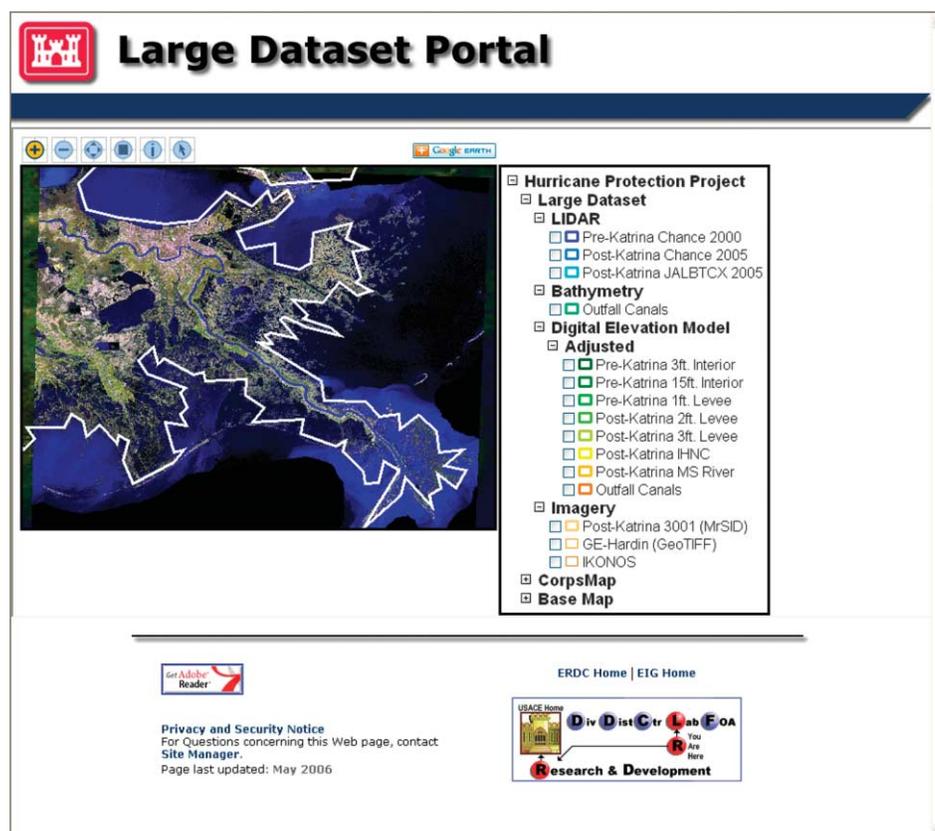
Figure 5 | Thin-client web application of the data browser.

computational models that reflect future and emerging directions in ecohydraulics, water resources, environmental, and sedimentation modeling. The list is not intended to be exhaustive and will be updated as new models are being developed or updated. Table 1 also identifies the corresponding framework that will support the model.

MODEL INTEGRATION EXAMPLES

ERDC is actively pursuing a number of model integration systems and methodologies. Some already exist while others are in the planning stage. This section will document these active areas of research for model integration.

HEC-WAT

The Hydrologic Engineering Center's Watershed Analysis Tool (WAT) (Dunn *et al.* 2005) is designed to help coordinate project studies to provide greater consistency

between multidisciplinary teams. Although models are not necessarily linked together at runtime, the data for each HEC-WAT execution is located together in a directory so that individual model data can be easily located. Included in this directory is a common Data Storage System (HEC-DSS) (Charley & Pabst 1998) time series data file that contains all DSS time series data for all models, alternatives, and time windows. When a model is executed, the input and output are passed to the model along with the time window and other information. The way this information is passed will depend on how the model is executed in HEC-WAT. Usually it will be supplied as a script or an input file.

A model created outside of HEC-WAT can be incorporated into the system by copying its contents into the HEC-WAT directory and rewiring the DSS file and common data to the common directory. The first step in the execution of a HEC-WAT simulation is to define the dimensions of the watershed along with background layers

Table 1 | Component tools of the USACE integrated modeling and assessment suite

Modeling domain	Model name	Framework	Reference
Watershed/ecological	EDYS	GIS	Childress *et al.* (1999)
Watershed/hydrologic	HMS	GIS/XMS/WAT	Feldman (2000)
Watershed/hydrologic	GSSHA	XMS	Downer & Ogden (2004)
Watershed/hydrologic	ADH	XMS	Stockstill & Berger (2001)
Riverine/ecological	EFM	WAT	Dunn & Hickey (2003)
Riverine/hydraulic	RAS	GIS/XMS/WAT	Brunner (2002)
Riverine/hydraulic	ADH	XMS	Stockstill & Berger (2001)
Riverine/environmental	CEQUAL-W2	XMS	Cole & Wells (2002)
Estuarine/hydraulic	ADH	XMS	Allard *et al.* (2002)
Estuarine/hydraulic	ADCIRC	XMS	Atkinson *et al.* (2004)
Nearshore/hydraulic	ADCIRC	XMS	Atkinson *et al.* (2004)
Nearshore/hydraulic	STWAVE	XMS	Smith *et al.* (2001)
Ocean/hydraulic	ADCIRC	XMS	Atkinson *et al.* (2004)
Groundwater/hydraulics	ADH	XMS	Kavanagh *et al.* (2002)
Groundwater/hydraulics	WASH123D	XMS	Yeh *et al.* (2005)
Reservoir/environmental	CEQUAL-W2	XMS	Cole & Wells (2002)

such as gage layers, stream alignment layers, and surface TINs or digital elevation models (DEMs). Next, a stream alignment must either be created or imported. The third step is to define common computation points (CCP), which are locations where one model transfers data to another model or where results are needed for further analysis.

As different alternatives are being created, a table with each alternative is constructed that shows required models and events to be analyzed for completion of the project. The table can be used to view the status of the data development (not started, in progress, or completed) and to show which data sets have been run and which need to be rerun due to changes in upstream conditions. A default program sequence is also generated in the table to show the order that models can be run within HEC-WAT. For example, HEC-HMS may be needed to generate information used by HEC-ResSim, which provides data for HEC-RAS.

Because HEC-WAT does not link models during runtime, there is no automated model control. However, the table keeps track of what models and events still need to be performed for the completion of a particular project. Some models do require input from other models in order to be run, but it is not automated.

In HEC-WAT, each computational model saves model output as a HEC-DSS file, and the next logical model reads data from the file in the same format. Currently, all models must read and write HEC-DSS files in order to be compliant with the HEC-WAT system. Model input files
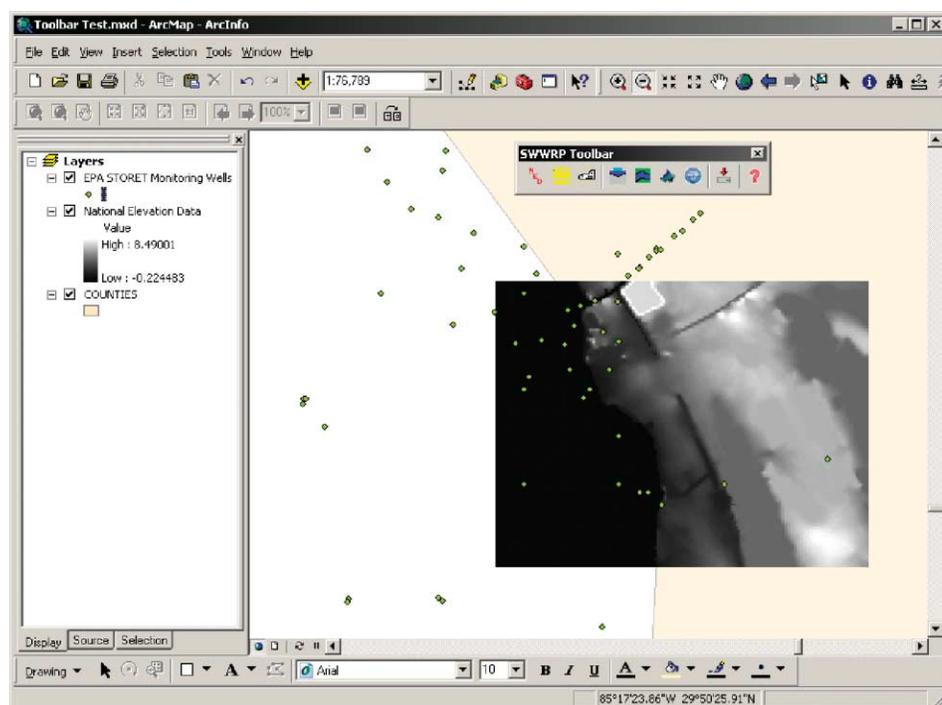
Figure 6 │ ArcView extension of the data browser.

are created in each individual modeling system. For instance riverine models are set up using the Riverine Analysis System (HEC-RAS), watershed models in the Hydrologic Modeling System (HEC-WMS), and reservoir models in the HEC-ResSim system. Once each individual model is created, they are then combined into a unified system within the HEC-WAT user interface (Figure 7). Each model is represented by a specific icon, and HEC-WAT is used to create the logical connections between each model. HEC-WAT then manages these connections by keeping track of all the data file locations, the data types of the linkages between models, the logical input and output points within the model setup, and the physical locations where models are being linked. Model execution is managed by HEC-WAT using the master control execution methodology described above. The HEC-WAT user sets the model execution sequence by selecting program order from the HEC-WAT interface. Once execution of all the models is complete, analysis of modeling results is done within the model interface designed for each specific model.

SMS steering module

The Surface-water Modeling System Steering Module (SMS-SM) (Zundel *et al.* 2002; http://chl.erdc.usace.army. mil/Media/3/9/7/SMS8-Fact%20Sheet.pdf) was designed to facilitate a specific link between two models within the SMS interface. These models communicate through files. The data from an output file is read and used as the input for another model input file.

Because the models are already integrated into the SMS interface, each model can be set up with its necessary parameters in SMS. The Steering Module can then be implemented to control the linkage between the models. The Steering Module has a dialog, shown in Figure 8, in which the user can specify how long one model will run before reading in information supplied by the other model. After setting these parameters, the process is completely automated, with the Steering Module controlling the model runs.

One model runs for the specified amount of time before needing information from the other model. The first model breaks and the second model runs, supplying the
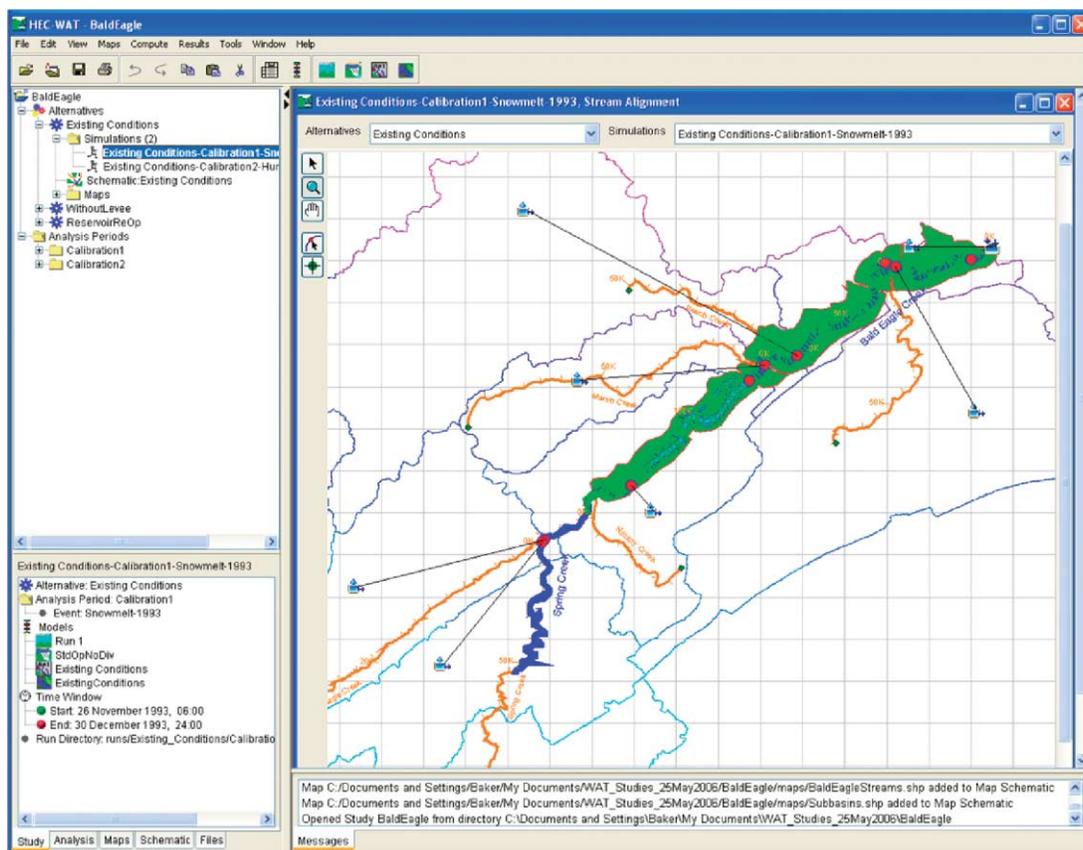
Figure 7 | HEC-WAT user interface.

information necessary for the first model to continue running. This may happen several times during a simulation. SMS runs on Windows XP and 2000.

SMS-SM is also a master control type of model integration system. It is a model coupling system that allows for modeling scenarios that require feedback between models during execution. This type of model coupling is required to compute tidally dominated riverine scenarios, the effect of wind and waves on near surface ocean currents, or the impact of sedimentation processes on a hydrodynamic solution. Similarly to HEC-WAT, each computational model is first set up using the appropriate model set-up tools. For instance, in a wind/wave/circulation scenario, both ADCIRC and STWAVE are set up using the Surface Water Modeling System. Because SMS-SM can handle multidimensional models, a different data transfer system is required. SMS-SM uses a mapping object that defines where data will be transferred between models. These objects are

set up before coupled execution begins using one of the XMS user interfaces. The mapping objects handle any spatial interpolation that may be required due to model domain differences. If modeling scenarios use models that operate at different temporal scales, SMS-SM can perform temporal aggregation and interpolation. Currently SMS-SM handles only a two-model coupled scenario.

For inclusion within SMS-SM, each computational model must be modified to allow execution to stop and start with changing input conditions. SMS-SM proceeds by first starting execution of the first model. When it has computed the assigned number of time-steps, it stops and writes output to a file formatted in the XMDF format. SMS-SM then executes a mapping function to convert data from one spatial and temporal form to that required by the other model. SMS-SM then starts execution of the second model. Execution is transferred back and forth, with interpolation being performed each time, until the entire coupled scenario is completed.
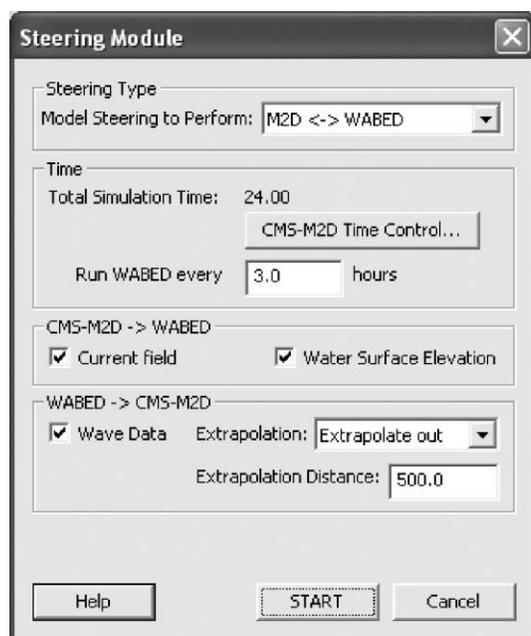
Figure 8 | SMS Steering Module dialog.

SUMMARY

The need for developing an integrated modeling and assessment infrastructure stems from the fact that at present – and probably even more so in the future – environmental modeling and water resources management require a holistic approach, combining a wide range of processes and supporting an even wider range of management decisions. Hence hydroinformatics tools and technologies are being developed with the objective to provide the following:

- A basic set of tools and utilities that can employed to solve problems of differing scope, scale, and technical sophistication across the life cycle of feasibility, planning, engineering design, construction, operations, and maintenance,
- The ability to input data once and reuse it many times seamlessly across the project lifecycle,
- The basic assessment and evaluation platforms through which new science can be fielded, by which alternatives can easily be evaluated, and from which decisions can be made.

Each of these tools has its own specialized set of capabilities (e.g. operates on portions of an ecohydraulic problem at fundamental time and space scales). However,

stakeholders (users, customers, proponents, decision makers) are concerned primarily about one thing: getting effective solutions to their problems. The developments described in this paper reflect on the required technological infrastructure to empower productive use of hydroinformatics modeling and assessment tools within the ecohydraulics and water resources management decision-making context.

REFERENCES

Allard, R., Barron, C., Blain, C. A., Hogan, P., Keen, T., Smedstad, L., Wallcraft, A., Berger, C., Howington, S., Smith, J., Signell, R., Bettencourt, M. & Cobb, M. 2002 High fidelity simulations of littoral environments. Presented at the *2002 Department of Defense High Performance Computing Modernization Program Users Conference, Austin, TX, June 10–14*, 2002. Naval Research Laboratory, Oceanography Division, Stennis Space Center, MS, USA.

Atkinson, J. H., Westerink, J. J. & Luettich, R. A. 2004 Two dimensional dispersion analysis of finite element approximations to the shallow water equations. *Int. J. Numer. Meth. Fluids* **45** (7), 715–749.

Blind, M. & Gregersen, J. B. 2005 Towards an Open Modelling Interface (OpenMI): the HarmonIT project. *Adv. Geosci* **4**, 69–74. SRef-ID: 1680-7359/adgeo/2005-4-69. European Geosciences Union.

Brunner, G. W. 2002 *HEC-RAS, River Analysis Systems Hydraulic Reference Manual*. US Army Corps of Engineers, Hydrologic Engineering Center (HEC), Report CPD-69.

Campbell, A. P. & Hummel, J. R. 1998 *The Dynamic Information Architecture System: An Advanced Simulation Framework For Military And Civilian Applications*. Argonne National Laboratory, Argonne. Illinois, USA. http://www.dis.anl.gov/DIAS/papers/SCS/SCS.html.

Charley, W. & Pabst, A. 1998 Hydrologic Engineering Center NEXGEN Data Storage System (HEC-DSS). In *Proceedings of the First Federal Interagency Hydrologic Modeling Conference: Bridging the Gap Between Technology and Implementation of Surface Water Quantity and Quality Models in the Next Century, Las Vegas, NV, April 19–23*, Interagency Advisory Committee on Water Data, Hydrology Subcommittee. Reston, VA, USA. pp. 5–33.

Childress, W. M., Price, D. L., Coldren, C. L. & Mclendon, D. T. 1999 *A Functional Description of the Ecological Dynamics Simulation (EDYS) Model, with Applications for Army and Other Federal Land Managers*. CERL Technical Report 99/55, US Army Corps of Engineering, June.

Cole, T. & Wells, S. 2002 *CEQUAL-W2: A Two-dimensional, Laterally Averaged, Hydrodynamic and Water Quality Model, Version 3.1*. Instruction Report EL-02-1. US Army Corps of Engineers.

Collins, N., Theurich, G., DeLuca, C., Suarez, M., Trayanov, A., Balaji, V., Li, P., Yang, W., Hill, C. & da Silva, A. 2005 Design and implementation of components in the Earth System

Modeling Framework. *Int. J. High Perf. Comput. Appl.* Fall/ Winter. **19** (3), 341–350.

Downer, C. W. & Ogden, F. L. 2004 GSSHA: model to simulate diverse stream flow producing processes. *J. Hydrol. Engng.* **9** (3), 161–174.

Dunn, C. N. P., Brunner, G. W. P. & Harris, J. 2005 Software integration for watershed studies: Hydrologic Engineering Center's Watershed Analysis Tool (HEC-WAT). In *Proceedings of the 2005 World Water and Environmental Resources Congress. EWRI 2005: Impacts of Global Climate Change, Anchorage, AK, 15–19 May*. Raymond Walton, ed. American Society of Civil Engineers. Reston, VA, Available from ASCE on CD-ROM.

Dunn, C. & Hickey, J. T. 2003 The Ecosystem Functions Model (EFM): a tool for restoration planning. In *Proceedings of the World Water and Environmental Resources Congress 2003, Philadelphia, PA, June 23–26* (eds. in P. Bizier & P. DeBarry). American Society of Civil Engineers. Reston, VA, Available from ASCE on CD-ROM.

Feldman, A. D. 2000 *Hydrologic Modeling System: HEC-HMS*, Technical Reference Manual. US Army Corps of Engineers, Hydrologic Engineering Center (HEC), Report CPD-74B.

Fine, S., Howard, S., Eyth, A., Herington, D. & Castleton, K. 2002 The EPA Multimedia Integrated Modeling System Software Suite. In *Presented at Second Federal Interagency Hydrologic Modeling Conference, Las Vegas, NV, July 28–August 1*. PowerPoint presentation.

Jones, N. L., Jones, R. D., Butler, C. D. & Wallace, R. M. 2004 A generic format for multi-dimensional models. In *Proceedings of the Groundwater Symposium, EWRI '04 World Water & Environmental Resources Congress, Salt Lake City, UT, June 27–July 1*. American Society of Civil Engineers. Reston, VA, Available from ASCE on CD-ROM.

Kavanagh, K. R., Kelley, C. T., Berger, R. C., Hallberg, J. P. & Howington, S. E. 2002 Nonsmooth nonlinearities and temporal integration of Richards' equation. In *Computational Methods in Water Resources XIV*, vol. 2 (eds. Hassanizadeh, M.Schotting, R. J. Gray, W. G. & Pinder, G. F.). Elsevier. Amsterdam pp. 947–954.

Leavesley, G. H., Restrepo, P. J., Markstrom, S. L., Dixon, M. & Stannard, L. G. 1996 *Modular Modeling System (MMS)– User's Manual*, US Geological Survey. Open-File Report pp. 96–151.

Smith, J. M., Sherlock, A. R. & Resio, D. T. 2001 *STWAVE: Steady-state Spectral Wave Model User's Manual for STWAVE Version 3.0*. CE-ERDC/CHL IR-01-1. US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory. Vicksburg, MS.

Stockstill, R. L. & Berger, R. C. 2001 *Simulation of flow in hydraulic structures using ADH*. In *Proceedings of the World Water & Environmental Resources Congress, Orlando FL, May 20–24*. American Society of Civil Engineers. Reston, VA, Available from ASCE on CD-ROM.

Whelan, G., Buck, J. W., Hoopes, B. L., Castleton, K. J., Pelton, M. A. & Gelston, G. M. 1998 Concepts associated with a framework for risk analysis in multimedia environmental systems. In *Proceedings of the American Nuclear Society – Topical Meeting on Risk-Based Performance Assessment and Decision Making*. American Nuclear Society. La Grange Park, IL, PNNL-SA-30475.

Yeh, G. T., Huang, G., Cheng, H. P., Zhang, F., Lin, H. C., Edris, E. & Richards, D. 2005 *A First-Principle, Physics-Based Watershed Model: WASH123D, Watershed Models* (eds. Singh V. P. & Frevert D. K.). CRC Press. Boca Raton, FL, chapter 9.

Zundel, A. K., Cialone, M. A. & Moreland, T. J. 2002 *SMS Steering Module for Coupling Waves and Currents, 1. ADCIRC and STWAVE*. Coastal and Hydraulics Engineering Technical Note ERDC/CHL CHETN-IV-41. US Army Engineer Research and Development Center. Vicksburg, MS.